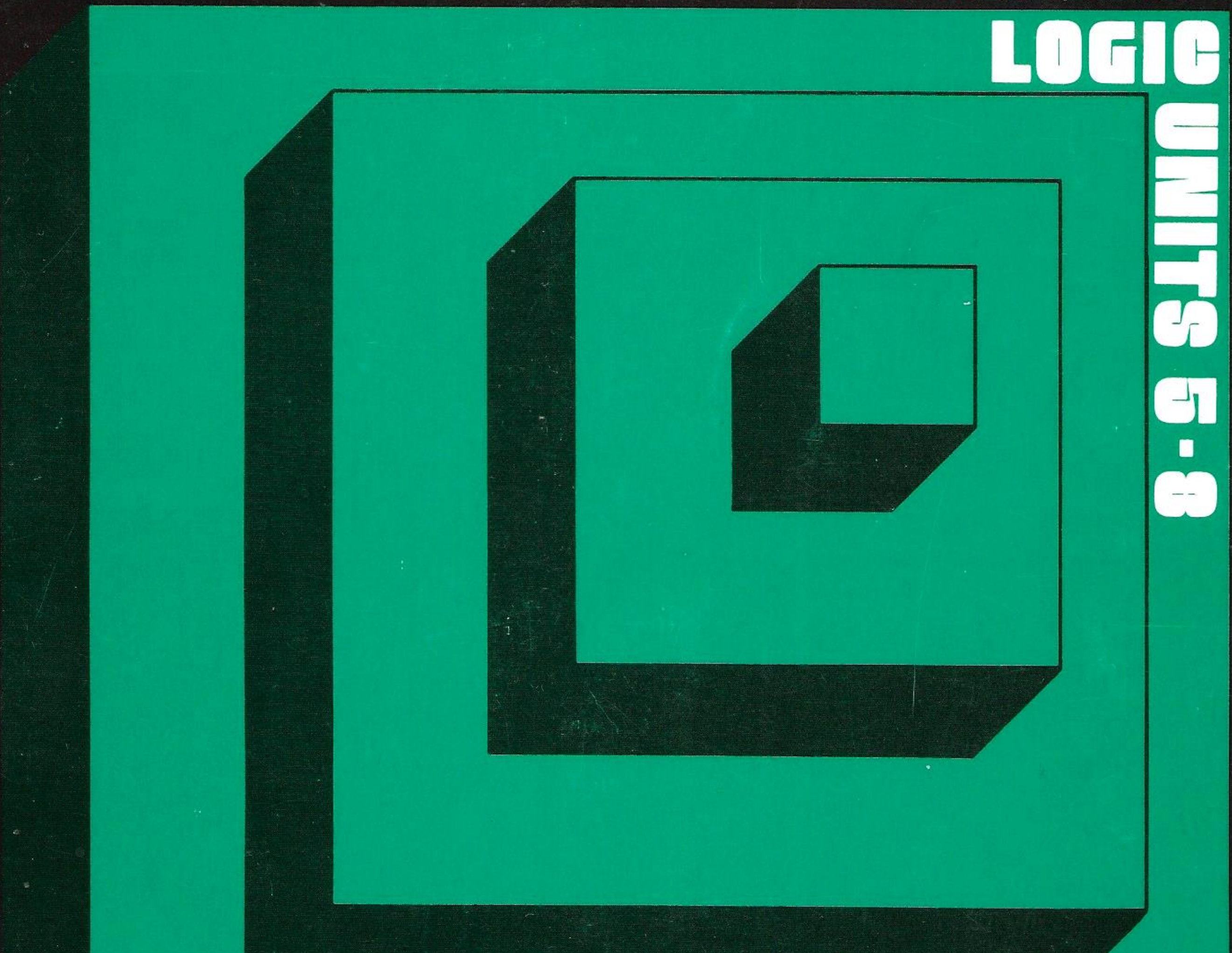
M335 L 5-8

THE OPEN UNIVERSITY
Mathematics: A Third Level Course
M335 Studies in Pure Mathematics

STUDIES IN PURE MATHEMATICS

LOGIC

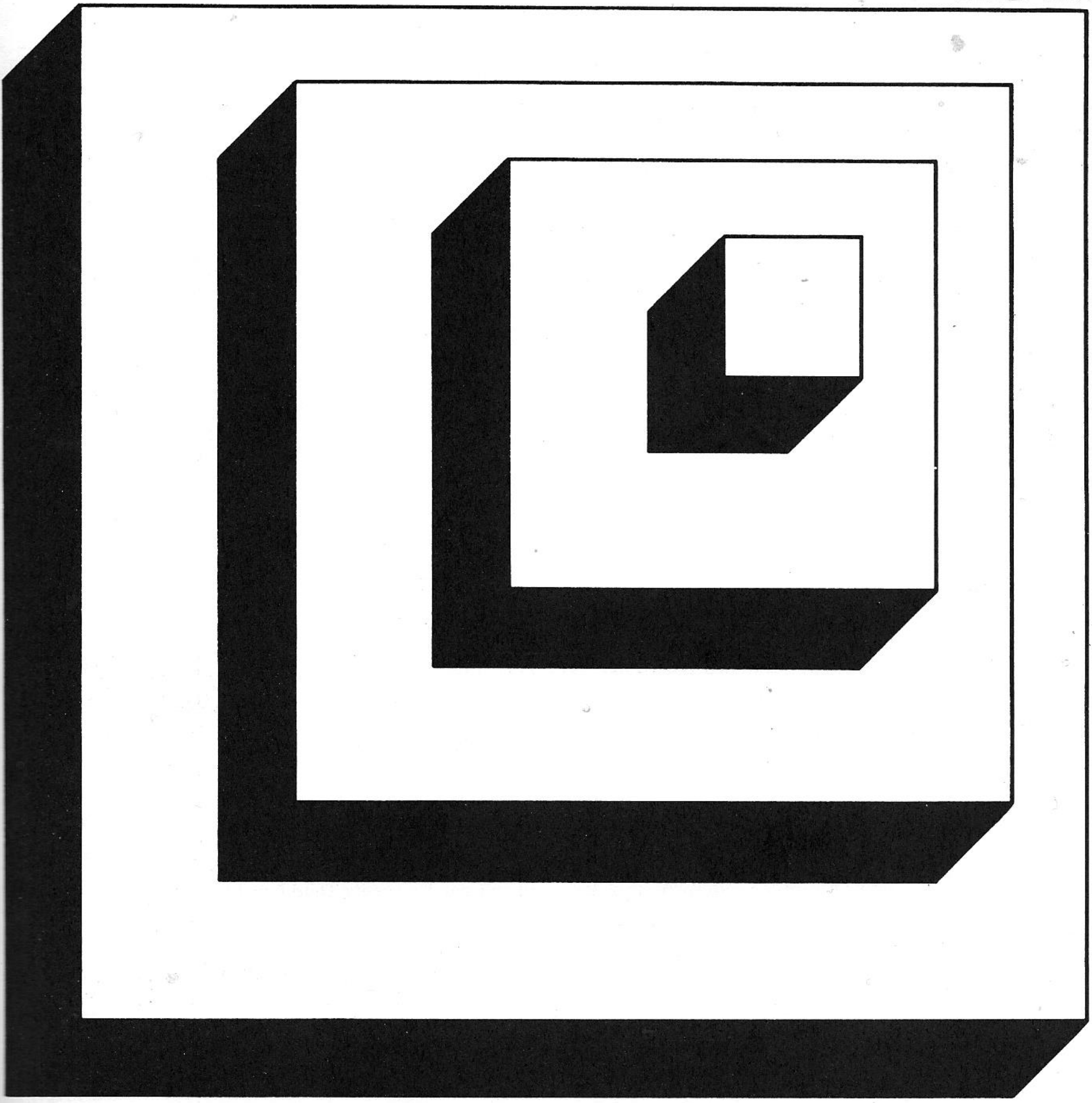
UNITS 5-8



THE OPEN UNIVERSITY
Mathematics: A Third Level Course
M335 Studies in Pure Mathematics
Prepared by the Course Team

LOGIC

UNITS 5-8



To do this option it is essential that you have a copy of the set book: *Computability and Logic* by G. Boolos and R. Jeffrey (Cambridge University Press, 1980, 2nd edition). In the text, this book will be referred to as **BJ**. You will need access to an audio-cassette player as there are audio-tapes accompanying the units.

It is hoped that you will find time to attempt all the Problems in the units. If, however, you are pressed for time, make sure to read our solutions to the problems, as these often contain teaching material. The objectives of each unit are stated purely in terms of what material might be tested in the assessment of the course, and can be found as the last section of each unit.

M335 Logic Option Course Team

Andrew Brown
Graham Flegg
Derek Goldrei
Alison Hughes
Ros Porter
Alan Slomson (University of Leeds)
Alex Wilkie (University of Oxford)
Hossein Zand

The Open University, Walton Hall, Milton Keynes

First published 1981. Reprinted 1984, 1988, 1993, 1997, 1999

Copyright © 1981 The Open University

All rights reserved. No part of this work may be reproduced, in any form, by mimeograph or any other means, without permission in writing from the publisher.

Designed by the Graphic Design Group of the Open University.

Produced in Great Britain by Page Bros, Norwich.

ISBN 0 335 14013 0

This text forms part of the correspondence element of an Open University Third Level Course.

For general availability of supporting material referred to in this text, please write to Open University Educational Enterprises Limited, 12 Cofferidge Close, Stony Stratford, Milton Keynes, MK11 1BY, Great Britain.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Milton Keynes MK7 6AB.

1.6

Contents

Unit 5 Formal Systems

5.0 Introduction **6**
5.1 The Language of Formal Number Theory **7**
5.2 Formal Proof **17**
5.3 Logical Consequence and Truth Tables **21**
5.4 The Tautology Rule and Conditional Proof **31**
5.5 Objectives of Unit 5 **35**
5.6 Appendix: A Turing Machine for Terms **36**

Unit 6 Quantifier Logic

6.0 Introduction **40**
6.1 Free and Bound Variables **41**
6.2 The Universal Quantifier Elimination Rule **44**
6.3 The Universal Quantifier Introduction Rule **48**
6.4 The Existential Quantifier Introduction Rule **52**
6.5 The Existential Hypothesis Rule **54**
6.6 Identity Rules **59**
6.7 The Correctness and Adequacy of Our System **62**
6.8 Objectives of Unit 6 **63**
6.9 Appendix: Guidelines for Finding Formal Proofs **63**

Unit 7 Formal Number Theory

7.0 Introduction **66**
7.1 An Axiom System for Number Theory **66**
7.2 Proofs in the System *Q* **69**
7.3 Non-standard Interpretations of *Q* **74**
7.4 The Representability Theorem **78**
7.5 Objectives of Unit 7 **83**

Unit 8 Gödel's Incompleteness Theorems

8.0 Introduction **86**
8.1 Gödel Numbers **86**
8.2 Gödel's Diagonal Lemma **88**
8.3 The Answer to Leibniz's Question **92**
8.4 Gödel's First Incompleteness Theorem **95**
8.5 Peano Arithmetic **98**
8.6 An Answer to Hilbert's Question **99**
8.7 In Conclusion **102**
8.8 Objectives of Unit 8 **103**

UNIT 5
FORMAL SYSTEMS

5.0 Introduction

The Role of Units 5 and 6

Boolos and Jeffrey assume that their readers are already familiar with formal logic. In Chapter 9 they remind their readers about the basic ideas of first-order logic (which is the only logical system we shall be considering in this course) and in Chapters 10–13 they prove and discuss some interesting theorems about it. *In this course we do not assume that you have met formal logic before, so Units 5 and 6 contain an introduction starting from scratch.* Our work is directed towards Gödel's Incompleteness Theorems, and in order to be able to reach them we shall leave out Chapters 10–13, though we shall mention briefly (without proofs) the chief facts about first-order logic contained in those chapters.

What is a Formal System?

Mathematics is usually written in a natural language such as English, augmented with some special symbols, like $\dfrac{dy}{dx}$ and $\displaystyle\int_a^b$. Natural languages have great advantages but from our point of view they also have the defects of vagueness and ambiguity. In a *formal language*, however, we first state explicitly what the basic symbols of the language are, these being the analogues of the letters of the alphabet and the punctuation marks of written English, and then state precise rules for determining which sequences of these basic symbols count as expressions of the language. (This already contrasts with a natural language—could you give a precise rule for deciding which sequences of letters of the alphabet are English words?!)

These rules are such that there is an algorithm for determining which sequences of our basic symbols are expressions of the language. Thus we could construct a Turing machine with the following property. If an arbitrary string of basic symbols is written on the tape (with the symbols of the string written on successive tape squares) and the machine is started scanning the left-most basic symbol, then the machine will halt with a single 1 on the tape if the original string was an expression of the language, and halt with a blank tape otherwise. Of course, for there to be such a Turing machine, our formal language can have only finitely many basic symbols, and this will indeed be the case.

Usually, when carrying out mathematical arguments we do not state explicitly which logical principles we are using. In a *formal system*, the rules used in carrying out deductions are stated explicitly. Again they have an algorithmic character, so we could devise a Turing machine which would check whether a purported proof of the system really was a proof.

What is the Use of a Formal System?

Recall that in the historical background to Gödel's theorems (as described on the first cassette tape), we asked you to keep in mind two questions at which this course is directed. Each of these questions provides a motivation for the setting up of a formal system for number theory.

(i) *Leibniz's question:* is there an algorithm for deciding which statements of number theory are true?

To discover whether there is an algorithm for deciding which statements of number theory are true, clearly we need first to be able to express the statements of number theory in a form to which algorithms are applicable, that is, as strings of symbols. Also, if we had no algorithm to decide which such strings were expressions, it would seem that we would have no hope of deciding which expressed true statements. Thus we must have a formal language for number theory in the sense described above.

(ii) *Hilbert's question:* can the consistency of number theory be proved using only non-dubious principles of finitary reasoning?

Hilbert wanted a completely clear and *finitary* proof that number theory is consistent: that is, a proof, involving only the mechanical manipulation of finite strings of symbols, that one could never prove contradictory assertions about natural numbers. Thus one clearly requires, in order to make sense of this programme, a precise notion and theory of proofs.

Once we have established a formal system for number theory, we can exploit the fact that the proofs of the system are obtained by manipulation of finite strings of symbols according to algorithmic rules, to try to show, using simple reasoning, that these manipulations cannot lead to a contradiction. That is, we can carry out Hilbert's programme of trying to give a finitary consistency proof for number theory.

We must, therefore, begin the description of our formal system, a task which will take us up to *Unit 7*. Our first requirement is a language.

5.1 The Language of Formal Number Theory

In this section we give a list of the basic symbols of the language, and the rules for determining which sequences of these symbols are expressions of the language.

The Basic Symbols

The first move in trying to decide which symbols we shall need might be to pick up a book on number theory and make a list of all the symbols it uses. This would, however, lead to an extravagantly long list. Our study of Turing machines has shown already the advantages of keeping things simple, and we approach the choice of symbols for number theory in the same spirit. We shall, in fact, discover that we can manage with remarkably few basic symbols.

Consider, for example, how we might express Euclid's Theorem that there are infinitely many prime numbers. We can do this by saying that, given any number, there is always a prime number which is at least as large as the given number. Thus, using the symbol $\forall$ to mean 'for all', the symbol $\exists$ to mean 'there is' (or 'there exists') and the symbol & to mean 'and', we get

$$\forall x\ \exists y\ (\text{‘}y \text{ is prime'}\ \&\ x \leq y).$$

With the aim of getting rid of the English phrase 'is prime', we note that the idea of a prime number can be expressed in terms of divisibility. So if we use $z|y$ for 'z divides y', then we can express 'y is prime' as

$$\forall z\ (z|y \ \rightarrow\ (z = y \text{ or } z = 1))\ \&\ y \neq 1,$$

where we have used the symbol → to mean 'implies' (i.e. 'if... then'). We ought to say here that we are regarding all our variables (z, y, x etc.) as referring to natural numbers 0, 1, 2, 3,..., rather than to, say, integers or real numbers. Thus the first part of the above expression (before the symbol &) is literally translated as: 'for all natural numbers z, if z divides y then z equals y or z equals 1'. Clearly a natural number y will satisfy this statement if and only if y is prime or y equals 1. The number 1 is not usually regarded as a prime number, so we have added the symbols & $y \neq 1$ at the end of the expression.

But divisibility is defined in terms of multiplication, so that $\exists t\ (z \cdot t = y)$ expresses $z|y$. Also '$x \leq y$' can be expressed in terms of addition, namely by $\exists s\ (s + x = y)$. (Notice that the expression $\exists s\ (s + x = y)$ is equivalent to $x \leq y$ only because we are regarding the variable s as referring to *natural* numbers. If we are referring to integers say, then $\exists s\ (s + x = y)$ would certainly *not* express $x \leq y$. For example, $\exists s\ (s + 3 = 2)$ is true if we are referring to integers but false if we are referring to natural numbers.) So, using $\vee$ to stand for 'or', and putting all this together, we see that Euclid's Theorem can be expressed by

$$\forall x\ \exists y(\forall z\ (\exists t\ (z \cdot t = y) \rightarrow (z = y\ \vee\ z = 1))\ \&\ y \neq 1\ \&\ \exists s\ (s + x = y))$$

Thus it turns out that Euclid's Theorem can be expressed using a very limited stock of symbols. Some of these symbols, for example the symbol &, would be needed in any mathematical theory. These are called *logical symbols.* $\forall$, $\rightarrow$ and $=$ are clearly also logical symbols. Any mathematical theory also uses *variables* to refer to the objects under consideration in that theory. Since we shall be interested in number theory, we intend the variables that we use to refer to natural numbers, although in fact, the logical rules that we eventually introduce for carrying out deductions will be valid no matter to which set of objects the variables are referring.

Other symbols, such as $\mathbf{+}$, are very special to number theory, and we call these *arithmetic symbols.* The arithmetic symbols include symbols for addition and multiplication, and apparatus to provide symbols for all natural numbers. It might be thought that this would involve having infinitely many different symbols. However, just as we use monadic notation for simplicity to represent numbers on the tapes of Turing machines, when it comes to formal number theory, we find we can make do with a symbol $\mathbf{0}$ to denote zero, and a symbol to denote the operation 'add one'. (Thus we can denote, for example, the number two by $\mathbf{0}''$.) As with the variables, because we shall be interested in number theory, we intend the arithmetic symbols to represent the usual arithmetic operations on the set of natural numbers, e.g. $\mathbf{+}$ for addition. However, the formal system we shall describe will handle these symbols in a purely mechanical way that does not necessarily take into account their intended interpretation. To distinguish these symbols from their interpretations, we shall use a heavier print for the symbols, e.g. $\mathbf{+}$ rather than $+$.

DEFINITION The *basic symbols* of the language of formal number theory are:

(a) *Logical symbols* These come in five categories.

- (i) *Connectives:* $- \quad \& \quad \vee \quad \rightarrow \quad \longleftrightarrow$ (called respectively negation conjunction, disjunction, implication and bi-implication).
- (ii) *Quantifiers:* $\forall\ \exists$ (known respectively as the universal quantifier and the existential quantifier).
- (iii) *Identity:* $=$
- (iv) *Punctuation symbols:* ( )
- (v) *Variables:* $x_0, x_1, x_2, \ldots$

(b) *Arithmetic symbols* $\mathbf{0}\ \mathbf{'}\ \mathbf{+}\ \boldsymbol{\cdot}$ (known respectively as the zero symbol, the successor symbol, the addition symbol and the multiplication symbol). ▲

Notice that the variables form an infinite list. If we don't like the idea of having infinitely many different symbols, we can regard the variables as made up from the two symbols x and $,$, so that, for example, x_4 can be regarded as an abbreviation for '$x_{,,,,}$'.

We have already mentioned the intended meanings of the connectives & ('and'), $\vee$ ('or'), $\longrightarrow$ ('implies' or equivalently 'if... then') and the quantifiers $\forall$ ('for all') and $\exists$ ('there is' or 'there exists'). It might help you to know at this stage that the intended meaning of the connective $-$ is 'not', and of $\longleftrightarrow$ is 'if and only if'. We shall discuss the interpretations of all symbols more fully later in the course.

Before giving the precise rules for deciding which sequences of these symbols are to be regarded as proper expressions of our formal language, let us take some examples of informal number theoretic statements and formulate them using only the symbols above. We hope that this will give you some idea of how mathematically rich these few symbols are.

XAMPLE 1 'Every number is either even or odd.'

Firstly note that 'x_0 is even' may be written as $\exists x_1\ x_0 = 2 \cdot x_1$. However, 2 is not a basic symbol, but as mentioned above, we can write 2 as $\mathbf{0}''$, which contains only basic symbols. Thus 'x_0 is even' becomes $\exists x_1\ x_0 = (\mathbf{0}'' \cdot x_1)$. Similarly, '$x_0$ is odd' becomes $\exists x_1\ x_0 = (2x_1 + 1)$ or, in terms of basic symbols, $\exists x_1\ x_0 = ((\mathbf{0}'' \cdot x_1) + \mathbf{0}')$.

Thus 'every number is either even or odd' and can be written as

$$\forall x_0\ (\exists x_1\ x_0 = (\mathbf{0}'' \cdot x_1) \vee \exists x_1\ x_0 = ((\mathbf{0}'' \cdot x_1) + \mathbf{0}')),$$

and this is just a string of basic symbols.

XAMPLE 2 'Every number has a square root.'

A first reaction here is perhaps 'but this is false!'. However, the truth or falsity of these examples is not our present concern. For the moment we just want to show that this informal statement can be expressed with a string of basic symbols. Indeed, we shall wish to be able to express statements of number theory without knowing whether they are true or false—clearly Leibniz's question would be meaningless unless we could do this.

'x_0 has a square root' can be written as $\exists x_1\ x_1^2 = x_0$. Now x_1^2 is not made up of basic symbols, but $x_1 \cdot x_1$ is, so we may write 'x_0 has a square root' as $\exists x_1\ (x_1 \cdot x_1) = x_0$; and hence the informal statement 'every number has a square root' becomes $\forall x_0\ \exists x_1\ (x_1 \cdot x_1) = x_0$ in our formal language. ☐

Formation Rules for Terms

The formation rules specify which sequences of the basic symbols are well-formed expressions of the language. The basic statements of the language consist of equations between numerical expressions (such as $x_0 = ((\mathbf{0}'' \cdot x_1) + \mathbf{0}')$ of Example 1 or $(x_1 \cdot x_1) = x_0$ of Example 2), and more complicated statements are built up from these using the connectives and quantifiers. So we begin with a precise definition of which sequences of symbols are well-formed numerical expressions. We call such expressions *terms*. The idea of the definition of a term is this. We first specify that the string consisting of the single symbol $\mathbf{0}$ is a term and that each string consisting of just a single variable (i.e. $x_0, x_1, \ldots$) is a term, and then state how more complicated terms may be built up from these terms by successive applications of addition, multiplication and the successor symbols. The key feature of this building process will be that it can be applied in reverse to *any* string of basic symbols and we will eventually arrive back at the symbol $\mathbf{0}$ and variables if and only if the original string was indeed a term. For example, this analysis applied to the string

$$((\mathbf{0}'' \cdot x_1) + \mathbf{0}')$$

will go as follows. The string $((\mathbf{0}'' \cdot x_1) + \mathbf{0}')$ is the result of applying the addition symbol $+$ to the strings $(\mathbf{0}'' \cdot x_1)$ and $\mathbf{0}'$. Now the string $\mathbf{0}'$ is the result of applying the successor symbol $'$ to the string $\mathbf{0}$, and this part of the analysis is complete. It remains to analyse the string $(\mathbf{0}'' \cdot x_1)$. This string is the result of applying the multiplication symbol $\cdot$ to the strings $\mathbf{0}''$ and x_1. Since x_1 is a variable, it remains only to analyse the string $\mathbf{0}''$, which is the result of applying the successor symbol to the string $\mathbf{0}'$, which is itself the result of applying the successor symbol to the string $\mathbf{0}$. Since in all cases we have reached the symbol $\mathbf{0}$ or a variable, the analysis is complete and we deduce that the original string $((\mathbf{0}''\cdot x_1) + \mathbf{0}')$ is indeed a term.

Notice that in this analysis we eliminated the symbol $+$ before the symbol $\cdot$ This order of doing things is of course dictated by the position of the parentheses. The string $(\mathbf{0}''\cdot(x_1 + \mathbf{0}'))$ would be analysed in a different way from the string $((\mathbf{0}''\cdot x_1) + \mathbf{0}')$, and obviously has a different intended meaning. So our definition must be specific on the positioning of brackets. Here, then, is our precise definition of a term.

DEFINITION

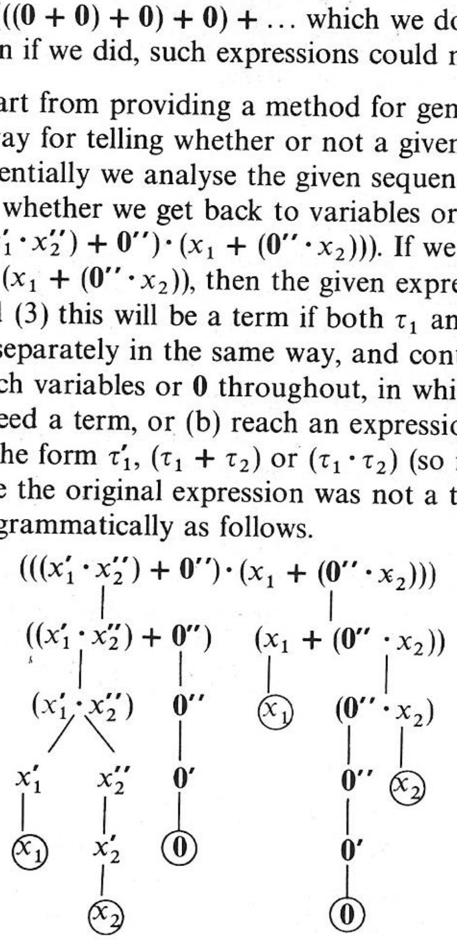
(1) $\mathbf{0}$ and each variable by itself are terms.

(2) If τ_1, τ_2 are terms then so also are τ_1', $(\tau_1 + \tau_2)$ and $(\tau_1\cdot\tau_2)$.

(3) A sequence of symbols is a *term* only if it follows from a finite number of applications of (1) and (2) that it is a term. ▲

Clause (2) of the definition says that, for example, if we place the symbol $+$ between two expressions which are terms, and a pair of brackets round the outside, then the resulting expression is also a term. Thus, as the expressions $\mathbf{0}$ and x_3 are both terms by clause (1), $(\mathbf{0} + x_3)$ is a term by clause (2). We can, of course, apply clause (2) again to $(\mathbf{0} + x_3)$ and, say, x_3 (since we now know that both these expressions are terms) to deduce that $(\mathbf{0} + x_3)'$ and $((\mathbf{0} + x_3)\cdot x_3)$ are both terms, and so on to obtain more complicated terms. Clause (3) now tells us that an expression is a term only provided it can be obtained in this way. This clause is needed to ensure that gobbledegook expressions such as $+\mathbf{0}x_0($ do not get regarded as terms. We have insisted on a *finite* number of applications of clauses (1) and (2) since an infinite number of such applications would clearly result in an expression of infinite length (e.g. we could obtain $\ldots(((\mathbf{0} + \mathbf{0}) + \mathbf{0}) + \mathbf{0}) + \ldots$ which we do not regard as a string of symbols; and even if we did, such expressions could not be handled by a Turing machine).

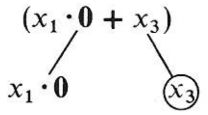
Apart from providing a method for generating terms, the definition also suggests a way for telling whether or not a given sequence of basic symbols is a term. Essentially we analyse the given sequence by applying clause (2) in reverse to see whether we get back to variables or $\mathbf{0}$. For example, consider the expression $(((x_1'\cdot x_2'') + \mathbf{0}'')\cdot(x_1 + (\mathbf{0}''\cdot x_2)))$. If we write τ_1 for $((x_1'\cdot x_2'') + \mathbf{0}'')$ and τ_2 for $(x_1 + (\mathbf{0}''\cdot x_2))$, then the given expression is $(\tau_1\cdot\tau_2)$. Hence by clauses (2) and (3) this will be a term if both τ_1 and τ_2 are terms. We now analyse τ_1 and τ_2 separately in the same way, and continue this process until we either (a) reach variables or $\mathbf{0}$ throughout, in which case the original expression was indeed a term, or (b) reach an expression which is neither a variable nor $\mathbf{0}$ nor of the form τ_1', $(\tau_1 + \tau_2)$ or $(\tau_1\cdot\tau_2)$ (so no further analysis is possible), in which case the original expression was not a term. This analysis may be represented diagrammatically as follows.



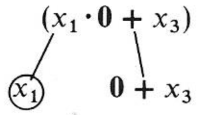
$(((x_1'\cdot x_2'') + \mathbf{0}'')\cdot(x_1 + (\mathbf{0}''\cdot x_2)))$
$((x_1'\cdot x_2'') + \mathbf{0}'')$
$(x_1 + (\mathbf{0}''\cdot x_2))$
$(x_1'\cdot x_2'')$
$\mathbf{0}''$
x_1
$(\mathbf{0}''\cdot x_2)$
x_1'
x_2''
$\mathbf{0}'$
$\mathbf{0}''$
x_2
x_1
x_2'
$\mathbf{0}$
$\mathbf{0}'$
x_2
$\mathbf{0}$

The circled expressions are all terms by clause (1) of our definition. The diagram shows how we can build up more complicated terms using clause (2). Every expression which is a term can be analysed in this way to show that it is a term. This sort of analysis will also reveal which expressions are not terms. Take, for instance, the expression $+\mathbf{0}x_0($. This expression is clearly neither a variable by itself nor the symbol $\mathbf{0}$. The expression does not end in a $'$ so it is not of the form τ_1'; nor does it begin and end with appropriate brackets, so it is not of the form $(\tau_1 + \tau_2)$ or $(\tau_1 \cdot \tau_2)$. Thus it cannot be analysed by either clause (1) or clause (2) of our definition, so it is not a term.

A more subtle example of an expression that is not a term is $(x_1 \cdot \mathbf{0} + x_3)$. Our analysis of this could start as


$(x_1 \cdot \mathbf{0} + x_3)$
$x_1 \cdot \mathbf{0}$
x_3

or as

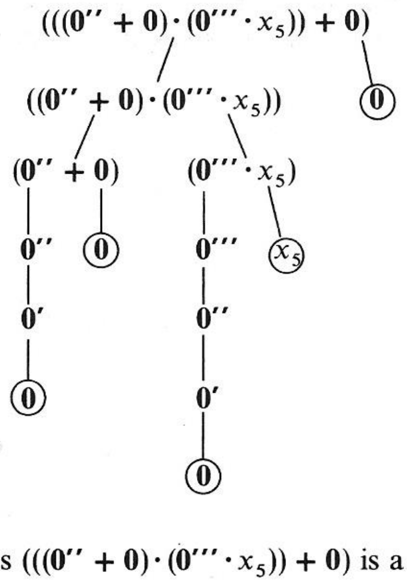

$(x_1 \cdot \mathbf{0} + x_3)$
x_1
$\mathbf{0} + x_3$

Either way we obtain an expression, $x_1 \cdot \mathbf{0}$ or $\mathbf{0} + x_3$, which is clearly neither a variable by itself nor the symbol $\mathbf{0}$—so we cannot analyse it further by clause (1)—and which, because of the lack of appropriate brackets, is not of the form τ_1', $(\tau_1 + \tau_2)$ or $(\tau_1 \cdot \tau_2)$—thus we cannot analyse it further by clause (2), and so it is not a term.

Essentially, the reason why we have formulated our definition of term so that the expression $(x_1 \cdot \mathbf{0} + x_3)$ is not a term, is because this expression is ambiguous: it could stand for $((x_1 \cdot \mathbf{0}) + x_3)$ or $(x_1 \cdot (\mathbf{0} + x_3))$. In everyday mathematics, one might have a convention that one of the symbols $+$, $\cdot$ has precedence over the other, to ensure that an expression like $(x_1 \cdot \mathbf{0} + x_3)$ is unambiguous. Although we could adopt such a convention in our formal system, we prefer to ensure this unambiguity by means of the use of the outer pairs of brackets in the expressions $(\tau_1 + \tau_2)$ and $(\tau_1 \cdot \tau_2)$ in clause (2).

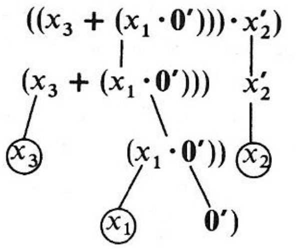
Here are some further examples of analyses of expressions to check whether or not they are terms.

EXAMPLE 3 $(((\mathbf{0}'' + \mathbf{0}) \cdot (\mathbf{0}''' \cdot x_5)) + \mathbf{0})$ is analysed as follows.


$(((\mathbf{0}'' + \mathbf{0}) \cdot (\mathbf{0}''' \cdot x_5)) + \mathbf{0})$
$((\mathbf{0}'' + \mathbf{0}) \cdot (\mathbf{0}''' \cdot x_5))$
$\mathbf{0}$
$(\mathbf{0}'' + \mathbf{0})$
$(\mathbf{0}''' \cdot x_5)$
$\mathbf{0}''$
$\mathbf{0}$
$\mathbf{0}'''$
x_5
$\mathbf{0}'$
$\mathbf{0}''$
$\mathbf{0}$
$\mathbf{0}'$
$\mathbf{0}$

Thus $(((\mathbf{0}'' + \mathbf{0}) \cdot (\mathbf{0}''' \cdot x_5)) + \mathbf{0})$ is a term.

EXAMPLE 4 $((x_3 + (x_1 \cdot \mathbf{0}'))) \cdot x_2')$ is analysed as follows:


$((x_3 + (x_1 \cdot \mathbf{0}'))) \cdot x_2')$
$(x_3 + (x_1 \cdot \mathbf{0}')))$
x_2'
x_3
$(x_1 \cdot \mathbf{0}'))$
x_2
x_1
$\mathbf{0}')$

Thus $((x_3 + (x_1 \cdot \mathbf{0}'))) \cdot x_2')$ is not a term since we arrive at an expression, namely $\mathbf{0}')$, which is plainly neither a variable nor $\mathbf{0}$, and cannot be analysed further by clause (2) of the definition.

PROBLEM 1 Analyse each of the following strings of basic symbols in the way described above and hence decide which are terms and which are not.

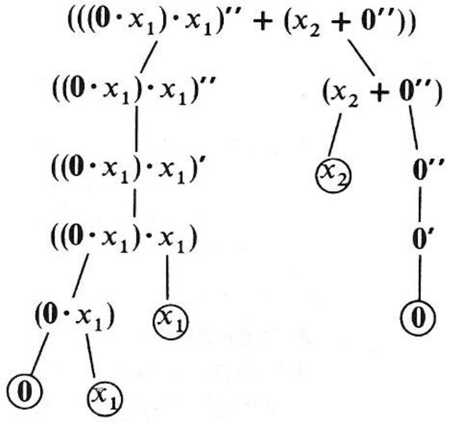
(i) $(((\mathbf{0} \cdot x_1) \cdot x_1)'' + (x_2 + \mathbf{0}''))$

(ii) $(((x_1 \cdot \mathbf{0}') \cdot \forall x_2) + (x_3'' \cdot (\mathbf{0})))$

(iii) $(x_6'' \cdot (x_9 + \mathbf{0}'''))'$

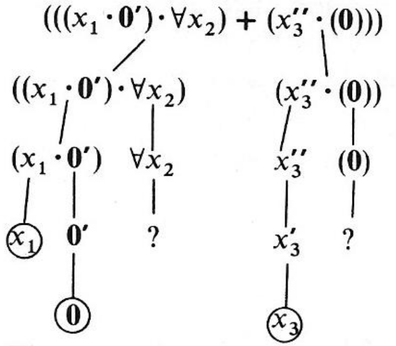
(iv) $(\mathbf{0} + + \mathbf{0})$

SOLUTION 1 (i)


$(((\mathbf{0} \cdot x_1) \cdot x_1)'' + (x_2 + \mathbf{0}''))$
$((\mathbf{0} \cdot x_1) \cdot x_1)''$
$(x_2 + \mathbf{0}'')$
$((\mathbf{0} \cdot x_1) \cdot x_1)'$
x_2
$\mathbf{0}''$
$((\mathbf{0} \cdot x_1) \cdot x_1)$
$\mathbf{0}'$
$(\mathbf{0} \cdot x_1)$
x_1
$\mathbf{0}$
$\mathbf{0}$
x_1

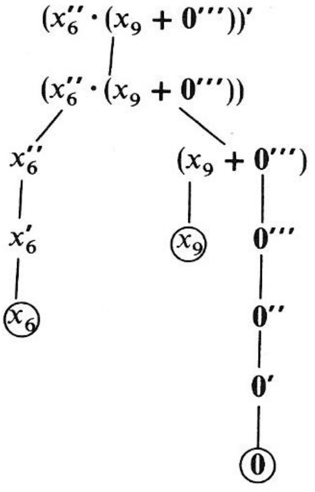
Hence (i) is a term.

(ii)


$(((x_1 \cdot \mathbf{0}') \cdot \forall x_2) + (x_3'' \cdot (\mathbf{0})))$
$((x_1 \cdot \mathbf{0}') \cdot \forall x_2)$
$(x_3'' \cdot (\mathbf{0}))$
$(x_1 \cdot \mathbf{0}')$
$\forall x_2$
x_3''
$(\mathbf{0})$
x_1
$\mathbf{0}'$
?
x_3'
?
$\mathbf{0}$
x_3

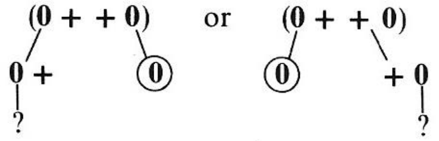
The expressions $\forall x_2$ and $(\mathbf{0})$ cannot be analysed further and are neither variables nor $\mathbf{0}$. Hence (ii) is not a term. (Of course it is easy to see that no term can contain the symbol $\forall$.)

(iii)


$(x_6'' \cdot (x_9 + \mathbf{0}'''))'$
$(x_6'' \cdot (x_9 + \mathbf{0}'''))$
x_6''
$(x_9 + \mathbf{0}''')$
x_6'
x_9
$\mathbf{0}'''$
x_6
$\mathbf{0}''$
$\mathbf{0}'$
$\mathbf{0}$

Hence (iii) is a term.

(iv) We could begin the analysis of this expression in two different ways:



but in either case we arrive at an expression, $\mathbf{0}$ + or + $\mathbf{0}$, which cannot be analysed using clause (2) of the definition and is neither the zero symbol nor a variable. Hence (iv) is not a term. ■

We hope that you are reasonably convinced that it is possible to devise a Turing machine that will recognize whether a given string of our basic symbols is or is not a term of our language. In case you are not sufficiently convinced, we have provided in the Appendix to this unit an outline of how such a machine could work. *However, this Appendix is purely optional reading and contains no assessable material: it should not be read if you are short of time.*

Formation Rules for Formulas

We are now in a position to make precise our formal equivalent of a mathematical statement, which we shall call a *formula* of our language. (This use of the word formula of course differs from the everyday mathematical use of the word.) Our basic building blocks will be the equivalents of statements asserting that two numerical expressions are equal. We have already defined the formal equivalent of numerical expressions, namely terms, and our formal language includes the identity symbol = with precisely the intention of representing equality. We call these basic building blocks *atomic formulas.*

EFINITION An expression is an *atomic formula* if and only if it has the form

$$\tau_1 = \tau_2$$

where τ_1 and τ_2 are terms. ▲

In other words, all the atomic formulas are obtained by taking two terms and placing the identity symbol between them. For example $x_0 = \mathbf{0}$, $x_1 = x_2$, $((x_1 \cdot x_2) + (x_1 \cdot x_3)) = (x_1 \cdot (x_2 + x_3))$, $x_2' = x_2$, are all atomic formulas since the expressions x_0, $\mathbf{0}$, x_1, x_2, x_2', $((x_1 \cdot x_2) + (x_1 \cdot x_3))$, and $(x_1 \cdot (x_2 + x_3))$ are all terms. $\mathbf{0} = \mathbf{0} \cdot \mathbf{0}$ is *not* an atomic formula since, although $\mathbf{0}$ is a term, $\mathbf{0} \cdot \mathbf{0}$ is not a term (as the brackets are missing). $x_1 = (\mathbf{0} \cdot x_2) = \mathbf{0}$ is *not* an atomic formula since it contains two occurrences of the identity symbol.

It should be fairly clear that there is an algorithm for determining whether or not a given expression is an atomic formula. Simply check that the given expression has exactly one occurrence of the identity symbol. If it doesn't, the expression is not an atomic formula; if it does, analyse the expressions on either side of the identity symbol separately to see if they are terms. If they both are terms, the original expression was an atomic formula; if at least one is not a term, the original expression was not an atomic formula.

ROBLEM 2 Which of the following are atomic formulas?

(i) $(x_1 + x_2)' = (x_1 + x_2')$

(ii) $(x_6'' \cdot (x_{17} + \mathbf{0}''))' = (\mathbf{0} + \mathbf{0})$

(iii) $(x_1 \cdot x_2) = (x_2 \cdot x_1)$

(iv) $x_4 =$

(v) $x_0 \cdot \mathbf{0} = \mathbf{0} \cdot x_0$

SOLUTION 2

(i) $(x_1 + x_2)'$ and $(x_1 + x_2')$ are both terms. So (i) is an atomic formula.

(ii) Similarly, this is an atomic formula.

(iii) Similarly, this is an atomic formula.

(iv) A blank space is not a term. So (iv) is not an atomic formula.

(v) Neither $x_0 \cdot \mathbf{0}$ nor $\mathbf{0} \cdot x_0$ are terms (as the brackets are missing), so (v) is not an atomic formula. ■

We now describe the set of all *formulas* of our formal language. The idea is that we allow as formulas exactly those strings of basic symbols that 'make mathematical sense' when we read ∀ as 'for all', ∃ as 'there exists', & as 'and', ∨ as 'or', – as 'not', → as 'implies' and ⟷ as 'if and only if'. However, as we have pointed out before, we must be precise enough for our definitions to be recognized by a machine, so this won't do as a definition. Instead, we build up more complicated formulas from the atomic formulas in a way analogous to that in which the terms are built up from **0** and the variables.

DEFINITION

(1) Each atomic formula is a formula.

(2) If ϕ, ψ are formulas and x is any variable, then $-\phi$, $(\phi \,\&\, \psi)$, $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$, $(\phi \longleftrightarrow \psi)$, $\forall x\ \phi$ and $\exists x\ \phi$ are also formulas.

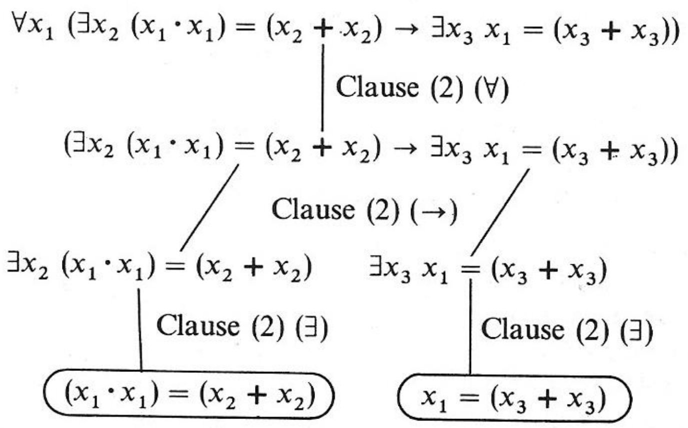
(3) A sequence of symbols is a *formula* only if it follows from a finite number of applications of (1) and (2) that it is a formula. ▲

Thus, for example, we have seen that $(x_1 + x_2)' = (x_1 + x_2')$ is an atomic formula, so it is a formula by clause (1). Since x_2 is a variable, it follows from clause (2) (about ∀) that $\forall x_2\ (x_1 + x_2)' = (x_1 + x_2')$ is a formula. Similarly, since x_1 is a variable it follows that $\forall x_1\ \forall x_2\ (x_1 + x_2)' = (x_1 + x_2')$ is a formula.

Because the definition has the same structure as the earlier definition of a term, the method of analysing sequences of symbols to determine whether or not they are formulas is similar to the analysis of terms we described earlier. Let us apply this method to show that the expression

$$\forall x_1\ (\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ x_1 = (x_3 + x_3))$$

is a formula.



$\forall x_1\ (\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ x_1 = (x_3 + x_3))$
Clause (2) (∀)
$(\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ x_1 = (x_3 + x_3))$
Clause (2) (→)
$\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2)$
$\exists x_3\ x_1 = (x_3 + x_3)$
Clause (2) (∃)
Clause (2) (∃)
$(x_1 \cdot x_1) = (x_2 + x_2)$
$x_1 = (x_3 + x_3)$

Notice that as long as we arrive at an expression containing either a quantifier or a connective we must be able to analyse further, until we reach an expression containing no quantifiers or connectives (those expressions circled in the above diagram). According to the definition, these must be atomic formulas (if the original string is a formula) and we already know how to check this. Here, we must check that the strings $(x_1 \cdot x_1) = (x_2 + x_2)$ and $x_1 = (x_3 + x_3)$ are atomic formulas, which we leave to the reader.

A proof that $\forall x_1\ (\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ \ x_1 = (x_3 + x_3))$ is a formula can be given by reading the diagram from bottom to top, as follows.

$(x_1 \cdot x_1) = (x_2 + x_2)$ and $x_1 = (x_3 + x_3)$ are atomic formulas (we assume you have checked this) and are therefore formulas by clause (1) of the definition.

Thus, since x_2 is a variable, $\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2)$ is a formula by clause (2). Similarly, $\exists x_3\ x_1 = (x_3 + x_3)$ is a formula.

Therefore $(\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ \ x_1 = (x_3 + x_3))$ is a formula by clause (2).

Thus, since x_1 is a variable,

$$\forall x_1\ (\exists x_2\ (x_1 \cdot x_1) = (x_2 + x_2) \rightarrow \exists x_3\ x_1 = (x_3 + x_3))$$

is a formula by clause (2).

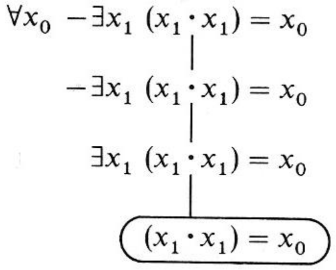
ROBLEM 3 Analyse each of the following expressions and hence decide which are formulas.

(i) $\forall x_0 - \exists x_1\ (x_1 \cdot x_1) = x_0$

(ii) $((\exists x_0\ x_1 = (x_0 + \mathbf{0}'') \lor x_1 = \mathbf{0}) \lor x_1 = \mathbf{0}')$

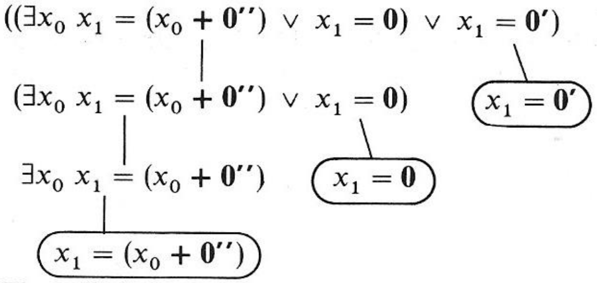
(iii) $(x_0 = \mathbf{0} \lor x_1 = \mathbf{0}\ \&\ x_2 = \mathbf{0})$

OLUTION 3 (i)


$\forall x_0 - \exists x_1\ (x_1 \cdot x_1) = x_0$
$- \exists x_1\ (x_1 \cdot x_1) = x_0$
$\exists x_1\ (x_1 \cdot x_1) = x_0$
$(x_1 \cdot x_1) = x_0$

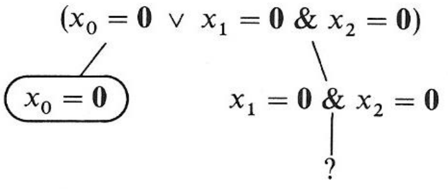
Thus (i) is a formula.

(ii)

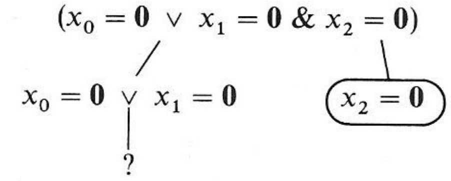

$((\exists x_0\ x_1 = (x_0 + \mathbf{0}'') \lor x_1 = \mathbf{0}) \lor x_1 = \mathbf{0}')$
$(\exists x_0\ x_1 = (x_0 + \mathbf{0}'') \lor x_1 = \mathbf{0})$
$x_1 = \mathbf{0}'$
$\exists x_0\ x_1 = (x_0 + \mathbf{0}'')$
$x_1 = \mathbf{0}$
$x_1 = (x_0 + \mathbf{0}'')$

Thus (ii) is a formula.

(iii) There are in fact two possible analyses here:


$(x_0 = \mathbf{0} \lor x_1 = \mathbf{0}\ \&\ x_2 = \mathbf{0})$
$x_0 = \mathbf{0}$
$x_1 = \mathbf{0}\ \&\ x_2 = \mathbf{0}$
?

or


$(x_0 = \mathbf{0} \lor x_1 = \mathbf{0}\ \&\ x_2 = \mathbf{0})$
$x_0 = \mathbf{0} \lor x_1 = \mathbf{0}$
$x_2 = \mathbf{0}$
?

In either case we arrive at an expression which is not atomic and cannot be analysed further. Thus (iii) is not a formula. In the first case, the expression $x_1 = \mathbf{0}\ \&\ x_2 = \mathbf{0}$ contains the connective & so certainly cannot be atomic. Also, it cannot be split using clause (2) (in reverse) since it is not in any of the required forms—essentially, the brackets are missing. Similarly, for $x_0 = \mathbf{0} \lor x_1 = \mathbf{0}$ in the second case. ■

As before, we hope that you are convinced that there is a Turing machine that can recognize whether or not a string of symbols written on its tape is a formula of our formal language. (If you are interested, an algorithm can be devised using a very similar technique to that used in recognizing terms dealt with in the Appendix. If ϕ is a formula which is not an atomic formula, define the *leading connective occurrence* (*l.c.o.*) *in* ϕ to be the last occurrence of $\forall$, $\exists$, $-$, $\&$, $\vee$, $\rightarrow$ or $\leftrightarrow$ used in building up ϕ using clause (2) in the definition of a formula. Then a Turing machine can recognize the l.c.o. in a formula ϕ using the 'counting brackets' argument, just as we recognize the l.s.o. in a term.)

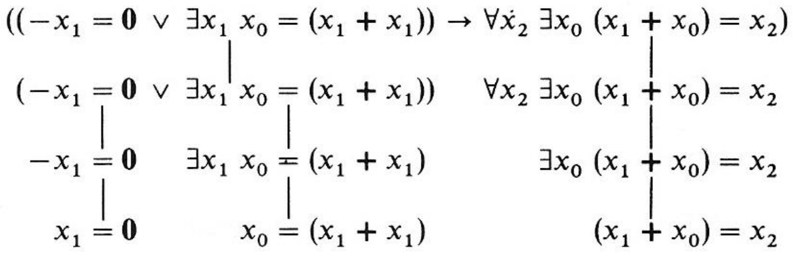
For technical reasons that will become apparent later, we now introduce the idea of a *subformula.*

DEFINITION The formulas that occur in the analysis of a given formula ϕ are called the *subformulas* of ϕ. We count ϕ as one of its own subformulas. ▲

EXAMPLE 5 Let us find the subformulas of the formula

$$((-x_1 = \mathbf{0} \vee \exists x_1\, x_0 = (x_1 + x_1)) \rightarrow \forall x_2\, \exists x_0\, (x_1 + x_0) = x_2).$$

Effectively we have to produce the analysis for this formula as if we were checking whether it is a formula. We obtain the following.


$((-x_1 = \mathbf{0} \vee \exists x_1\, x_0 = (x_1 + x_1)) \rightarrow \forall x_2\, \exists x_0\, (x_1 + x_0) = x_2)$
$(-x_1 = \mathbf{0} \vee \exists x_1\, x_0 = (x_1 + x_1))$
$\forall x_2\, \exists x_0\, (x_1 + x_0) = x_2$
$-x_1 = \mathbf{0}$
$\exists x_1\, x_0 = (x_1 + x_1)$
$\exists x_0\, (x_1 + x_0) = x_2$
$x_1 = \mathbf{0}$
$x_0 = (x_1 + x_1)$
$(x_1 + x_0) = x_2$

The subformulas of the formula are precisely the formulas obtained at each stage of this analysis. Thus the formula has the following nine subformulas.

(i) $x_1 = \mathbf{0}$
(ii) $-x_1 = \mathbf{0}$
(iii) $x_0 = (x_1 + x_1)$
(iv) $(x_1 + x_0) = x_2$
(v) $\exists x_1\, x_0 = (x_1 + x_1)$
(vi) $\exists x_0\, (x_1 + x_0) = x_2$
(vii) $(-x_1 = \mathbf{0} \vee \exists x_1\, x_0 = (x_1 + x_1))$
(viii) $\forall x_2\, \exists x_0\, (x_1 + x_0) = x_2$
(ix) $((-x_1 = \mathbf{0} \vee \exists x_1\, x_0 = (x_1 + x_1)) \rightarrow \forall x_2\, \exists x_0\, (x_1 + x_0) = x_2)$

PROBLEM 4 List the subformulas of the following formulas.

(i) $\forall x_0\, -\exists x_1\, (x_1 \cdot x_1) = x_0$ (This is the formula analysed in Problem 3(i).)
(ii) $((\exists x_0\, x_1 = (x_0 + \mathbf{0}'') \vee x_1 = \mathbf{0}) \vee x_1 = \mathbf{0}')$ (This is the formula analysed in Problem 3(ii).)
(iii) $x_1 = ((x_2 \cdot x_3) + \mathbf{0})$

SOLUTION 4 (i) From the analysis in Solution 3(i), we see that the subformulas are

(a) $(x_1 \cdot x_1) = x_0$
(b) $\exists x_1\, (x_1 \cdot x_1) = x_0$
(c) $-\exists x_1\, (x_1 \cdot x_1) = x_0$
(d) $\forall x_0\, -\exists x_1\, (x_1 \cdot x_1) = x_0$

(ii) From the analysis in Solution 3(ii), we see that the subformulas are

(a) $x_1 = (x_0 + \mathbf{0}'')$
(b) $x_1 = \mathbf{0}$
(c) $x_1 = \mathbf{0}'$
(d) $\exists x_0\, x_1 = (x_0 + \mathbf{0}'')$
(e) $(\exists x_0\, x_1 = (x_0 + \mathbf{0}'') \vee x_1 = \mathbf{0})$
(f) $((\exists x_0\, x_1 = (x_0 + \mathbf{0}'') \vee x_1 = \mathbf{0}) \vee x_1 = \mathbf{0}')$

(iii) Since this formula contains no occurrences of connectives or quantifiers, the analysis just looks like

$$x_1 = ((x_2 \cdot x_3) + \mathbf{0})$$

(i.e. the formula is atomic). Thus there is only one subformula, namely

$$x_1 = ((x_2 \cdot x_3) + \mathbf{0}). \blacksquare$$

As before, we hope you are convinced that a Turing machine could be devised which recognized subformulas of a given formula. We give no details of such a machine here.

5.2 Formal Proof

'Humpty Dumpty looked doubtful. "I'd rather see that done on paper," he said.'

Through the Looking Glass, Lewis Carroll.

We have already remarked that whereas in everyday mathematics we do not state explicitly which logical principles we are using, in a *formal system* the rules that can be used in carrying out deductions are stated explicitly. Before we describe the rules that we are going to allow, we need to think about what sort of thing an informal mathematical proof is, the intention being that our precisely defined notion of *formal proof* should correspond as closely as possible to the informal notion used by mathematicians.

Since the informal notion is not absolutely precise, we cannot hope to give an exact definition. The best we can expect is a more-or-less accurate description:

> 'A mathematical proof is a finite sequence of mathematical assertions which forms a correct and convincing argument for the desired conclusion from stated assumptions.'

'Sequence' means that the assertions are in some definite order, and 'finite' means that the argument comes to an end. 'Correct' means that the assertions do really follow from the stated assumptions. It is because proofs are correct arguments that when we have proved something we know it is true. In the next sections we begin to make more precise what it means to say that one assertion really does follow from another.

The correctness of an argument is a matter of logical fact; in contrast, 'convincing' is a psychological notion. From our point of view this is unsatisfactory. We want it to be a matter of objective fact that a sequence of assertions counts as a formal proof. Indeed, we require that there is actually an algorithm for determining whether or not a sequence of formulas counts as a formal proof, and it will become evident that the rules for carrying out deductions, which we shall describe later, meet this requirement.

The rest of this unit and all of *Unit 6* are devoted to describing our formal system. As we want formal proofs to correspond to the proofs we use in everyday mathematics, let us look at such an everyday proof, and see how this can be reorganized to bring out its logical structure. The example we have chosen might be found in an elementary book on number theory.

Theorem If the square of a natural number is even, the number itself is even.

Proof Assume n^2 is even. Now assume n is odd. Then, by definition, there is a natural number k such that $n = 2k + 1$. Hence $n^2 = (2k + 1)^2$. But $(2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, so $n^2 = 2(2k^2 + 2k) + 1$, which shows that n^2 is odd, contradicting our initial assumption. Hence n cannot be odd, so n is even. Thus, if n^2 is even, n itself is even. ■

The above is a perfectly good and correct everyday proof. However, it is not in the best form for a machine to check that it is indeed correct. This is largely because the proof leaves unsaid much of the justification of the steps it uses. For instance, when we write '$n^2 = (2k + 1)^2 \ldots$ and $(2k + 1)^2 = 2(2k^2 + 2k) + 1$, so $n^2 = 2(2k^2 + 2k) + 1$' we assume the reader is familiar with the rule that allows us to deduce that $a = c$ from the equalities $a = b$ and $b = c$ between natural numbers a, b, and c. Of course, in everyday maths, it is perfectly reasonable to assume that a reader will take this rule for granted. But a machine cannot take it for granted: a machine needs the rule and the use of the rule to be spelt out.

Thus, as a first step in our description of a formal proof, let us rewrite the above in a different format. Firstly, we shall use a separate line for each assertion made in the argument (changing the English in places, for convenience). Secondly, we shall give each line a number corresponding to the place of the assertion on that line in the argument. This will help us to refer quickly to the lines of the argument. (This is, of course, quite common practice in mathematics, for instance when one wishes to refer to important equations.) Next, to the right of each assertion, we shall write down the justification for that assertion—as this justification will often be in terms of earlier assertions, our line numbering is particularly useful here. Our rewritten proof is as follows.

Second Version of the Proof

Line number	*Assertion made*	*Justification*
(1)	n^2 is even	Assume
(2)	n is odd	Assume
(3)	For some number k, $n = 2k + 1$	From (2), by definition of 'odd'
(4)	$n^2 = (2k + 1)^2$	From (3)
(5)	$(2k + 1)^2 = 4k^2 + 4k + 1$	Algebraic manipulations
(6)	$4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$	
(7)	$(2k + 1)^2 = 2(2k^2 + 2k) + 1$	From (5) and (6)
(8)	$n^2 = 2(2k^2 + 2k) + 1$	From (4) and (7)
(9)	For some number l, $n^2 = 2l + 1$	From (8), putting $l = 2k^2 + 2k$
(10)	n^2 is odd	From (9), by definition of 'odd'
(11)	n^2 is odd and n^2 is even	From (10) and (1)
(12)	n is not odd	From (11), 'proof by contradiction'
(13)	n is even	From (12)
(14)	n^2 is even implies n itself is even	(13) has been deduced from (1)

The way that our layout separates the assertions made from their justification, i.e. the rules used to arrive at one assertion from previous ones, is a feature we shall build into our idea of formal proof. Thus a formal proof will consist of a sequence of assertions, each of which is accompanied by a reference to the rule by which it was derived from previous assertions.

In our informal proof, it seems that we are using a large number of different rules. But, in fact, we are going to reduce these to nine different rules which we shall explain in this unit and *Unit 6*. Before we start looking at these rules, there is one simple but important refinement we should like to make to the format of our proof, as follows.

Most proofs, formal and informal, contain assertions that are in fact assumptions, like lines (1) and (2) in the proof above. Various conclusions are drawn by working from these assumptions: for instance, by line (11) we have concluded from our two assumptions that n^2 is both even and odd (giving us a contradiction, so that it cannot be the case that both assumptions are simultaneously true). It turns out to be very helpful to keep a record on each line of the proof of the assumptions from which we have derived that line. For instance, if we look at the justification of line (8), we see that it follows from lines (4) and (7). Now line (4) follows from line (3), which in turn follows from line (2); but line (2) is an assumption which depends on nothing earlier in the proof. Also line (7) follows from lines (5) and (6), which are justified by simple algebraic properties of the natural numbers (such as the distributive law, commutative law etc.). Let us just write AP for the algebraic properties used here. (They are not really our main concern at this stage, although in a deeper analysis of this proof one would have to investigate exactly which such properties are being used.)

Thus the assertion on line (8) depends ultimately on the assumption on line (2) and AP. We usually express this fact by saying that the *assumptions in force* on line (8) are line (2) and AP, or that line (8) *depends on the assumption* (2) and AP, and we are going to indicate this by writing '2, AP' to the left of the line number. If we do this for each line we obtain the following.

Third Version of the Proof

Assumption	*Line number*	*Assertion*	*Justification*
1	(1)	n^2 is even	Assume
2	(2)	n is odd	Assume
2	(3)	For some number k, $n = 2k + 1$	From (2), by definition of 'odd'
2	(4)	$n^2 = (2k + 1)^2$	From (3)
AP	(5)	$(2k + 1)^2 = 4k^2 + 4k + 1$	
AP	(6)	$4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$	Algebraic manipulations
AP	(7)	$(2k + 1)^2 = 2(2k^2 + 2k) + 1$	From (5) and (6)
2, AP	(8)	$n^2 = 2(2k^2 + 2k) + 1$	From (4) and (7)
2, AP	(9)	For some number l, $n^2 = 2l + 1$	From (8), putting $l = 2k^2 + 2k$
2, AP	(10)	n^2 is odd	From (9), by definition of 'odd'
1, 2, AP	(11)	n^2 is odd and n^2 is even	From (10) and (1)
1, AP	(12)	n is not odd	From (11), 'proof by contradiction'
1, AP	(13)	n is even	From (12)
AP	(14)	n^2 is even implies n itself is even	(13) has been deduced from (1)

We have written the number 1 at the extreme left of line (1) because the assertion 'n^2 is even' is an assumption, and so naturally the assumption in force on line (1) is just the assertion that occurs there. Similarly for line (2). Now, to find the assumption in force on line (3), we look at the justification for the assertion made there, where we see that line (3) follows from line (2). So line (3) depends on the same assumptions that line (2) depends on. Line (2) has the number 2 written at its extreme left, so we write the number 2 to the left of line (3). Similarly, line (4) follows from line (3) where the assumption in force is just 2, so we write 2 to the left of line (4). We carry on in this way, keeping track of assumptions, until we reach line (11). (We leave you to verify the numbers written to the left of lines (5) to (10).) Line (11) follows from lines (10) and (1), so we pool all assumptions in force on those lines to get 1, 2, AP to the left of line (11). Now, what has been proved at this stage in the argument? Our analysis reveals (on line (11)) that the assumptions 1, 2, AP lead to a contradictory statement, namely 'n^2 is even and n^2 is odd'. Hence not all of these assumptions can be simultaneously true. Thus if we assume that line (1) and AP (i.e. the algebraic properties used) are true, then line (2) must be false. Now, line (2) is the assertion 'n is odd', so with the assumption 1 and AP it follows that 'n is not odd', and this is precisely what we have indicated on line (12). The proof is essentially complete at line (13) since this line states that the assertion 'n is even' follows from assumption 1, i.e. 'n^2 is even' (also assuming AP). However, it is neater to have the actual statement of the theorem we are proving on the last line of the proof, so we incorporate assumption 1 into the assertion on line (14), leaving AP as the only assumption on the left of this line.

These last lines in the third version of the proof illustrate the advantage of keeping track of the assumptions in force on each line. For a casual glance at line (13) in our second version of the proof might suggest that we have proved the statement 'n is even', which of course we have not. We have proved only that 'n is even' *assuming* 'n^2 is even', and this is made perfectly clear in line (13) of the third version. Also, as AP remains an assumption on line (14), we have revealed that some algebraic properties of the natural numbers have been used in the original proof.

Now the third version of our proof is still essentially informal. However, our notion of formal proof will copy its layout, so that a *formal proof* will consist of a table with columns representing:

Assumption	*Line number*	*Assertion*	*Justification*

An *assertion* in a formal proof will be simply a formula of the formal language (as defined in the previous section). The justification will consist of a reference to the use of *one of nine rules* that we shall define. The rules must, of course, tell us when one formula may be derived from others, and on what assumptions this formula depends. From the last line of the table we can claim that the assertion on that line has been formally proved from the assumptions in force on that line.

In the rest of this unit and in *Unit 6* we shall look at the rules that can be used in formal proofs and how they are used.

We begin with the formal rule that corresponds to the one used in lines (1) and (2) of the informal proof above, that is, the rule that corresponds to making an assumption. For this reason, the first rule of deduction of our formal system is called the rule of assumptions.

RULE *The Rule of Assumptions* (*Ass*)
Any formula may be introduced on a line of a proof. The assumption in force on this line is the formula itself. ▲

We use the annotation Ass on the right of a line to indicate that the Rule of Assumptions has been used. Thus a line on which this rule is used will have the following form.

Assumption	*Line number*	*Formula*	*Justification*
k	(k)	ϕ	Ass

Notice that the assumption number is the same as the line number. That is because, as specified by the rule, the assumptions in force on a line where the Rule of Assumptions is used consist just of the formula introduced as an assumption on that line. We can now give you an example of a formal proof:

1	(1)	$\exists x_0\, x_0 = x_0$	Ass

This is a formal proof of the formula $\exists x_0\, x_0 = x_0$ from the assumption of the formula on line 1, i.e. from the formula $\exists x_0\, x_0 = x_0$. Thus we have a proof of $\exists x_0\, x_0 = x_0$ assuming $\exists x_0\, x_0 = x_0$.

You might be pleased to learn that more rules will be introduced soon! But do not undervalue the Rule of Assumptions: formal proofs, and informal ones like the proof of 'if n^2 is even then n is even', can rarely be produced without using this rule, for instance to get started.

5.3 Logical Consequence and Truth Tables

We begin this section with an important observation that may have occurred to you already. It is that in specifying the formation rules of our language we have at no stage needed to refer to the intended meaning of our symbols. (Examples 1 and 2 of Section 5.1 did so, but in all subsequent examples we were careful never to refer to 'what is meant' by the symbols or formulas.) It is because we are able to specify which sequences of symbols are well-formed formulas without referring to meanings that we are able to devise a Turing machine to test whether or not a sequence of symbols is, in fact, a formula. Turing machines operate by recognizing the shapes and arrangements of symbols, but they cannot understand the meaning of the expressions written on their tapes.

When it comes to specifying the rules of deduction of our formal system, however, the intended interpretation of our symbols becomes important. Thus to say, for example, that 'from ϕ and $(\phi \to \psi)$ we may conclude ψ' is a valid inference, depends very definitely on our interpretation of the connective $\to$.

Why do we say that the above inference is valid? It is because whenever ϕ and $(\phi \to \psi)$ are interpreted as being true, then ψ must also be true. At first sight it may seem strange to talk about formulas being interpreted as being true, but here it is important to remember that although we have devised our language in order to express statements of number theory, nothing we have said so far implies that our formulas must be interpreted in that way. A formula which is true when interpreted as being about numbers may be false when given other interpretations. For example

$$\forall x_0\, \forall x_1\, (x_0 \cdot x_1) = (x_1 \cdot x_0)$$

is true if interpreted as referring to multiplication of positive integers, but false if interpreted as being about matrix multiplication, since, for example,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}$$

but

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}.$$

Logically valid inferences, therefore, are those which do not depend on any special interpretation given to the arithmetical symbols, but whose validity derives solely from the meanings given to the logical symbols (i.e. the connectives and the quantifiers). The logician Gottlob Frege described the use of logic as 'a way that, disregarding the particular characteristics of objects, depends solely on those laws on which all knowledge rests'. The laws to which Frege refers are those which tell us which deductions are logically valid. A deduction is logically valid if the conclusion is a logical consequence of the premises. The next important definition makes clear what we mean by *logical consequence.*

DEFINITION The formula ψ is a *logical consequence* of the formulas $\phi_1, \ldots, \phi_k$ if every interpretation of the non-logical symbols under which all the formulas $\phi_1, \ldots, \phi_k$ are true also makes the formula ψ true. ▲

For this definition to be precise, we must, of course, explain what we mean by 'an interpretation of the non-logical symbols making a formula true'. In our language of formal number theory, the non-logical symbols are just the four arithmetical symbols (**0**, $'$, $+$ and $\cdot$) and an account of their possible interpretations will be given in *Unit* 7. However, we wish now to emphasize the difference between the arithmetical symbols and the logical symbols, in order to make the above definition clearer.

While the arithmetical symbols will have many interpretations (for example, possible interpretations of the symbol $+$ will include addition of natural numbers, addition of real numbers, etc.), the logical symbols will have a fixed interpretation. The reason for this is best explained by looking at examples from ordinary English. Consider the following three arguments.

(1) If he races on soft ground then my horse will win.
My horse is racing today and the ground is soft.
Therefore my horse will win today.

(2) He hit the ball with the spin and he hit it hard.
Therefore he hit the ball hard.

(3) All men are mortal.
Socrates is a man.
Therefore Socrates is mortal.

Obviously all three are valid arguments, but from where does this validity derive? Not, we claim, from the actual subject matter of the statements involved. No knowledge of the turf, cricket or Socrates is required to see that the above arguments are correct. Rather, their validity derives solely from the everyday use of the English 'connecting' words 'if... then...', 'and', 'all' etc. Thus the truth of the statement 'he hit the ball with the spin and he hit it hard' *guarantees* the truth of the statement 'he hit the ball hard' simply because of the meaning of the word 'and', and not because of the meanings of 'hit', 'ball' or 'spin'.

Of course, the statement 'he hit the ball with the spin and he hit it hard' may be false or inapplicable in a given situation, so we may deduce nothing from it concerning that particular situation. However, this does not invalidate the argument (2) above, as the only thing that is being said there is that

> *if* in a certain situation it is true that 'he hit the ball with the spin and he hit it hard'
>
> *then* it must be true that, in the same situation, 'he hit the ball hard'.

Indeed, if A and B represent two statements,

> *if* we are given that the statement (A and B) is true
>
> *then* B is forced to be true (as is A, of course).

We hope that this shows up the 'content-free' nature of argument (2) above. Similarly, (1) and (3) are just particular cases of generally valid arguments. In argument (1) we are using just the meaning of the connecting words 'if... then...':

> *whenever* (if A then B) and A are both true
>
> *then* B must be true,

no matter what meanings the statements A and B have. Similarly argument (3) has the form:

> *if* (all X are Y) and (Z is X) are both true,
>
> *then* (Z is Y) is true,

which is valid again not because of the subject matter (since there is none in this abstract version) but because of the meanings of 'all', 'is' and 'are'.

To return to our formal system, it is the logical symbols (connectives and quantifiers) that will play a role similar to the English connecting words discussed above, while the arithmetical symbols will express the 'subject matter'. Our definition of logical consequence is thus intended to capture arguments such as (1), (2) and (3) above, that is to say, arguments such that the truth of the *conclusion* (ψ in the definition) is guaranteed by the truth of the *premises* ($\phi_1, \ldots, \phi_k$) solely because of the (fixed) interpretation of the logical symbols, and does not depend upon any particular interpretation of the arithmetical symbols (the 'subject matter').

We now describe the intended interpretation of the logical symbols. The rest of this unit will deal with the connectives, while *Unit 6* will begin with a discussion of the quantifiers.

(a) Conjunction

We use the symbol & for conjunction. If ϕ, ψ are formulas, then their conjunction is $(\phi \,\&\, \psi)$. Clause 2 of our definition of a formula tells us that this is also a formula.

The intended interpretation of the symbol & is explained by saying that:

The formula $(\phi \,\&\, \psi)$ *is true if and only if both the formulas* ϕ *and* ψ *are true.*

Another way to explain this is by means of the following table:

ϕ	ψ	$(\phi \,\&\, \psi)$
true	true	true
true	false	false
false	true	false
false	false	false

We call this the *truth table* for &. 'True' and 'false' are called *truth values* because we can think of them as values attained by the formulas ϕ and ψ. The above table completely specifies the interpretation of & because we have specified the truth value attained by $(\phi \,\&\, \psi)$ for *all* possible combinations of truth values attained by ϕ and by ψ.

For technical reasons it is more convenient to use 1 and 0 as truth values. We use 1 for 'true' and 0 for 'false'. Also, we usually write truth tables in an abbreviated form, so that the truth table for & will usually be written in the following way.

$(\phi$	$\&$	$\psi)$
1	1	1
1	0	0
0	0	1
0	0	0

↑
the values of
$(\phi \& \psi)$

The symbol & corresponds to the informal use of 'and'.

(b) Negation

We use $-$ for negation. The negation of the formula ϕ is the formula $-\phi$. The intended interpretation of the symbol $-$ is explained by saying that:

The formula $-\phi$ *is true if and only if the formula* ϕ *is false.*

Thus the truth table for $-$ is:

ϕ	$-\phi$
1	0
0	1

or in abbreviated form:

$-$	ϕ
0	1
1	0

The symbol $-$ corresponds to the informal use of 'not'.

(c) Disjunction

We use $\vee$ for disjunction. The disjunction of the formulas ϕ and ψ is the formula $(\phi \vee \psi)$. The truth table for $\vee$ is

$(\phi$	$\vee$	$\psi)$
1	1	1
1	1	0
0	1	1
0	0	0

Thus the intended interpretation of $\vee$ is explained by saying that:

The formula $(\phi \vee \psi)$ *is true if and only if either* ϕ *is true or* ψ *is true or they are both true.*

In other words $(\phi \vee \psi)$ is false if and only if ϕ is false and ψ is false.

The symbol $\vee$ corresponds to the informal use of 'or' that suggests alternatives which are not necessarily mutually exclusive.

(d) Implication

We use → for implication. If ϕ, ψ are formulas, the implication with *antecedent* ϕ and *consequent* ψ is the formula $(\phi \rightarrow \psi)$. The truth table for → is

$(\phi$	$\to$	$\psi)$
1	1	1
1	0	0
0	1	1
0	1	0

The intended interpretation of → can thus be explained by saying that:

> *The formula* $(\phi \to \psi)$ *is true if and only if either* ϕ *is false or* (ϕ *is true and* ψ *is true*).

Perhaps it is clearer to say $(\phi \to \psi)$ is false if ϕ is true and ψ is false, and $(\phi \to \psi)$ is true otherwise. The formal symbol → is intended to correspond to the informal notion 'implies' or 'if... then...'. If it seems strange that we define $(\phi \to \psi)$ to be true when ϕ is false (regardless of the truth value of ψ), the following example should help to explain this.

Consider the statement 'for all integers n, if $n > 2$ then $n^2 > 4$'. This is typical of how the words 'if... then...' are used in mathematics and, of course, one immediately accepts this statement as true. But this forces us to accept the statements 'if $-3 > 2$ then $9 > 4$' (putting $n = -3$) and 'if $2 > 2$ then $4 > 4$' (putting $n = 2$) as true, and in both cases the antecedent is false. In the first case, the consequent ('$9 > 4$') is true and in the second case the consequent ('$4 > 4$') is false. This illustrates why we have defined the third and fourth lines in the above truth table as we have.

(e) Bi-implication

We use ⟷ for bi-implication.

The intended interpretation of ⟷ can be explained by saying that:

> *The formula* $(\phi \longleftrightarrow \psi)$ *is true if and only if* ϕ *and* ψ *have the same truth value.*

The truth table for ⟷ is

$(\phi$	$\longleftrightarrow$	$\psi)$
1	1	1
1	0	0
0	0	1
0	1	0

The symbol ⟷ corresponds to the informal notion of 'if and only if'.

Using the basic truth tables, we can work out truth tables for more complicated formulas built up using the connectives. Our method is as follows. Suppose we have a formula θ built up from some of its subformulas, $\phi_1, \ldots, \phi_k$, using only connectives. To get one line of the truth table for θ, we assign a truth value to each of $\phi_1, \ldots, \phi_k$, and progressively work out the truth values of all the larger subformulas of θ corresponding to this assignment of truth values by using the rules given in the tables in (a) to (e). The order in which we deal with the subformulas corresponds to the way in which θ is built up as a formula from $\phi_1, \ldots, \phi_k$. Finally, we arrive at a truth value for θ (which is by definition a subformula of itself). We then repeat this process for all the possible assignments of truth values to $\phi_1, \ldots, \phi_k$ to obtain the truth table.

EXAMPLE 1 Consider the formula $(-(\phi \vee \psi) \rightarrow (\phi \,\&\, \psi))$, where ϕ, ψ are any given formulas. (They will, of course, be subformulas of the main formula.) We can work out its truth table as follows:

ϕ	ψ	$(\phi \vee \psi)$	$-(\phi \vee \psi)$	$(\phi \,\&\, \psi)$	$(-(\phi \vee \psi) \rightarrow (\phi \,\&\, \psi))$
1	1	1	0	1	1
1	0	1	0	0	1
0	1	1	0	0	1
0	0	0	1	0	0

The headings of the columns correspond to the subformulas obtained when building up $(-(\phi \vee \psi) \rightarrow (\phi \,\&\, \psi))$ from ϕ and ψ. The first two columns list all possible combinations of truth values that can be attained by ϕ, ψ. The third column computes the corresponding truth values of $(\phi \vee \psi)$ according to the table in (c) above. In the fourth column we have applied the table in (b) to the truth values in the third column, which therefore gives us the corresponding truth values of $-(\phi \vee \psi)$. The fifth column is computed using the table in (a) applied to the first two columns. Finally, the sixth column is worked out by applying the table in (d) to the fourth and fifth columns, which therefore gives us the truth values of $(-(\phi \vee \psi) \rightarrow (\phi \,\&\, \psi))$, as required.

Thus the table tells us that, for example, when ϕ is true and ψ is false (i.e. the second row of truth values), then the entire formula is true. ☐

Note that on each line of the truth table in Example 1, the truth value of the formula is completely determined by the values given to ϕ and ψ and the rules in tables (a) to (e). This is guaranteed by the strict way in which we insisted on the brackets in our definition of a formula. If we had been sloppy in this definition, something like $(\phi \,\&\, \psi \vee \theta)$ would not have had a well-defined truth table, as the order of precedence between & and $\vee$ would affect the truth values taken by the 'formula'.

We could write down the truth table in Example 1 succinctly as follows:

$(-$	$(\phi$	$\vee$	$\psi)$	$\rightarrow$	$(\phi$	$\&$	$\psi))$
0	1	1	1	1	1	1	1
0	1	1	0	1	1	0	0
0	0	1	1	1	0	0	1
1	0	0	0	0	0	0	0

③ ① ② ① ④ ① ② ①

The circled numbers indicate the order in which the columns were filled in. They are not part of the truth table and would usually be left out. The values in column ④ tell us the truth value of the entire formula corresponding to the values of ϕ, ψ as given in the columns labelled ①.

The truth table above has 4 lines, corresponding to the $4 = 2^2$ possible combinations of truth values for the two subformulas ϕ and ψ. For a formula built up from 3 different subformulas using only the connectives, there would be $2^3 = 8$ possible combinations and hence the truth table would have 8 lines, and so on. (Note that the order in which one records the truth values of a formula corresponding to the different combinations of truth values of its subformulas in a truth table does not matter. All that matters is that all possible combinations are dealt with in the table.)

EXAMPLE 2 Let us calculate the truth values of the formula $((-\phi \mathbin{\&} \psi) \longleftrightarrow \chi)$ for all possible truth values of its subformulas ϕ, ψ and χ. These values may be read from the following truth table, which we leave you to check.

ϕ	ψ	χ	$-\phi$	$(-\phi \mathbin{\&} \psi)$	$((-\phi \mathbin{\&} \psi) \longleftrightarrow \chi)$
1	1	1	0	0	0
1	1	0	0	0	1
1	0	1	0	0	0
1	0	0	0	0	1
0	1	1	1	1	1
0	1	0	1	1	0
0	0	1	1	0	0
0	0	0	1	0	1

Or, in abbreviated form:

$((-$	ϕ	$\&$	$\psi)$	$\longleftrightarrow$	$\chi)$
0	1	0	1	0	1
0	1	0	1	1	0
0	1	0	0	0	1
0	1	0	0	1	0
1	0	1	1	1	1
1	0	1	1	0	0
1	0	0	0	0	1
1	0	0	0	1	0
②	①	③	①	④	①

PROBLEM 1 Work out the truth tables of the following formulas.

(i) $((\phi \mathbin{\&} \psi) \mathbin{\&} \chi)$

(ii) $((\phi \rightarrow \psi) \rightarrow \phi)$

(iii) $(\phi \longleftrightarrow \phi)$

(iv) $(-(\phi \vee \psi) \longleftrightarrow \chi)$

PROBLEM 2 Let θ denote the following formula:

$$(((\exists x_0\, (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}') \mathbin{\&} (\exists x_0\, (x_0 + x_1) = \mathbf{0} \vee \forall x_3\, x_1 = x_3)) \rightarrow (x_1 = \mathbf{0}' \vee \forall x_3\, x_1 = x_3)).$$

Let ϕ, ψ, χ denote the formulas $\exists x_0\, (x_0 + x_1) = \mathbf{0}$, $x_1 = \mathbf{0}'$, $\forall x_3\, x_1 = x_3$ respectively. Check that ϕ, ψ, χ are subformulas of θ, and that θ takes the truth value 1, no matter which truth values are given to ϕ, ψ and χ.

PROBLEM 3 Show that the formula θ given by

$$(\forall x_0\, x_0 = x_1 \rightarrow (-\exists x_2\, (x_2 + \mathbf{0}'') = x_1 \rightarrow (\exists x_2\, (x_2 + \mathbf{0}'') = x_1 \rightarrow x_3 = \mathbf{0}')))$$

has truth value 1, under all interpretations of the symbols.

(As you do not yet know how to find the truth value of formulas like $\forall x_0\, x_0 = x_1$ under an interpretation of the non-logical symbols, you will need to identify subformulas of θ such that θ is built up from the subformulas using only connectives, i.e. *not* using quantifiers, and then construct an appropriate truth table.)

SOLUTION 1 (i)

((ϕ	&	ψ)	&	χ)
1	1	1	1	1
1	1	1	0	0
1	0	0	0	1
1	0	0	0	0
0	0	1	0	1
0	0	1	0	0
0	0	0	0	1
0	0	0	0	0
①	②	①	③	①

(ii)

((ϕ	$\to$	ψ)	$\to$	ϕ)
1	1	1	1	1
1	0	0	1	1
0	1	1	0	0
0	1	0	0	0
①	②	①	③	①

(iii)

(ϕ	$\leftrightarrow$	ϕ)
1	1	1
0	1	0
①	②	①

(iv)

(−	(ϕ	$\vee$	ψ)	$\leftrightarrow$	χ)
0	1	1	1	0	1
0	1	1	1	1	0
0	1	1	0	0	1
0	1	1	0	1	0
0	0	1	1	0	1
0	0	1	1	1	0
1	0	0	0	1	1
1	0	0	0	0	0
③	①	②	①	④	①

■

SOLUTION 2 Either by using the tree analysis, or by inspection, we see that ϕ, ψ, χ are subformulas of θ and that θ may be written as

$$(((\phi \to \psi) \& (\phi \vee \chi)) \to (\psi \vee \chi)).$$

We thus have the following truth table for θ.

(((ϕ	$\to$	ψ)	&	(ϕ	$\vee$	χ))	$\to$	(ψ	$\vee$	χ))
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	0	1	1	1	0
1	0	0	0	1	1	1	1	0	1	1
1	0	0	0	1	1	0	1	0	0	0
0	1	1	1	0	1	1	1	1	1	1
0	1	1	0	0	0	0	1	1	1	0
0	1	0	1	0	1	1	1	0	1	1
0	1	0	0	0	0	0	1	0	0	0
①	②	①	③	①	②	①	④	①	②	①

From column 4 we see that θ takes the truth value 1 for all possible truth values of its subformulas ϕ, ψ and χ. ■

OLUTION 3 Let ϕ, ψ and χ denote the formulas $\forall x_0\ x_0 = x_1$, $\exists x_2\ (x_2 + \mathbf{0}'') = x_1$, and $x_3 = \mathbf{0}'$ respectively. Then ϕ, ψ and χ are all subformulas of θ, and θ is built up from ϕ, ψ and χ by use of only connectives. In fact, θ is then given by $(\phi \rightarrow (-\psi \rightarrow (\psi \rightarrow \chi)))$, and has the following truth table.

$(\phi$	$\rightarrow$	$(-$	ψ	$\rightarrow$	$(\psi$	$\rightarrow$	$\chi)))$
1	1	0	1	1	1	1	1
1	1	0	1	1	1	0	0
1	1	1	0	1	0	1	1
1	1	1	0	1	0	1	0
0	1	0	1	1	1	1	1
0	1	0	1	1	1	0	0
0	1	1	0	1	0	1	1
0	1	1	0	1	0	1	0

Thus θ is true under all interpretations. Note that the success of this method depends on an apt choice of subformulas of θ. For instance, if we had chosen ϕ^*, ψ^* and χ^* to be $\forall x_0\ x_0 = x_1$, $-\exists x_2\ (x_2 + \mathbf{0}'') = x_1$, and $(\exists x_2\ (x_2 + \mathbf{0}'') = x_1 \rightarrow x_3 = \mathbf{0}')$ respectively, then θ would be given by $(\phi^* \rightarrow (\psi^* \rightarrow \chi^*))$ and would have the following truth table.

$(\phi^*$	$\rightarrow$	$(\psi^*$	$\rightarrow$	$\chi^*))$
1	1	1	1	1
1	0	1	0	0
1	1	0	1	1
1	1	0	1	0
0	1	1	1	1
0	1	1	0	0
0	1	0	1	1
0	1	0	1	0

This would seem to suggest that θ is not true under all interpretations. However, this result occurs only because the subformulas chosen, ψ^* and χ^* in particular, are inter-related—in a sense, they are too large. In fact, the one line in the second table on which θ appears to obtain the truth value 0 cannot be achieved under any interpretation, as the truth values of ψ^* and χ^* are interdependent in such a way that χ^* can never take the value 0 when ψ^* takes the value 1 (as ψ^* is $-\psi$ and χ^* is $(\psi \rightarrow \chi)$). ■

The property of always taking the truth value 1, possessed by the formulas described in Problems 2 and 3 above, is of such importance to us in formulating our rules of proof that we introduce the following definition.

EFINITION Suppose ϕ is a formula with subformulas $\psi_1, \ldots, \psi_k$ such that ϕ may be written in terms of $\psi_1, \ldots, \psi_k$ using only the logical connectives. Suppose further that ϕ takes the truth value 1 (i.e. true) no matter what truth values are given to its subformulas $\psi_1, \ldots, \psi_k$. Then ϕ is called a *tautology*. ▲

For example Solution 2 above shows that the formula θ is a tautology. Also, Solution 1(iii) shows that the formula $(\phi \longleftrightarrow \phi)$ is a tautology.

As the remarks at the end of Solution 3 above show, if we do not make an appropriate choice of subformulas, we may fail to detect that a formula ϕ is a tautology when a different choice would have revealed this fact. The safest thing is to analyse ϕ into subformulas $\psi_1, \dots, \psi_k$ which are as simple as possible subject to the necessary condition that ϕ can be written in terms of $\psi_1, \dots, \psi_k$ using only the connectives (and thus not using the quantifiers). The resulting truth table will then tell us quite definitely whether or not ϕ is a tautology.

DEFINITION Now suppose $\phi_1, \dots, \phi_k$ are any formulas and ψ is any formula. We say that ψ is a *tautological consequence* of $\phi_1, \dots, \phi_k$ if the formula

$$((\dots((\phi_1 \,\&\, \phi_2) \,\&\, \phi_3) \dots \&\, \phi_k) \rightarrow \psi) \qquad (*)$$

is a tautology.

(We hope that it is clear what we mean by the formula (*). For $k = 1$, it is the formula $(\phi_1 \rightarrow \psi)$; for $k = 2$, $((\phi_1 \,\&\, \phi_2) \rightarrow \psi)$; for $k = 3$, $(((\phi_1 \,\&\, \phi_2) \,\&\, \phi_3) \rightarrow \psi)$, and so on.) ▲

For example, let ϕ_1 be $(\exists x_0\, (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}')$, let ϕ_2 be $(\exists x_0\, (x_0 + x_1) = \mathbf{0} \vee \forall x_3\, x_1 = x_3)$, and let ψ be $(x_1 = \mathbf{0}' \vee \forall x_3\, x_1 = x_3)$. Then Solution 2 above shows that $((\phi_1 \,\&\, \phi_2) \rightarrow \psi)$ is a tautology, so that $(x_1 = \mathbf{0}' \vee \forall x_3\, x_1 = x_3)$ is a tautological consequence of $(\exists x_0\, (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}')$, $(\exists x_0\, (x_0 + x_1) = \mathbf{0} \vee \forall x_3\, x_1 = x_3)$.

It will be important for later work that a Turing machine exists that can decide whether or not a given formula is a tautological consequence of a given set of formulas. Firstly, we hope that it is clear that (in principle, at least) one could programme a Turing machine to check whether a formula written on its tape is a tautology. (We essentially require a machine that computes the subformulas of a formula. It is then just a matter of assigning all possible truth values to subformulas and using the truth table method to check whether the original formula always takes the value 1.) We now programme a machine such that if a sequence $\phi_1, \dots, \phi_k, \psi$, is given as input, it writes out the formula $((\dots(\phi_1 \,\&\, \phi_2) \,\&\, \dots \&\, \phi_k) \rightarrow \psi)$ and then checks if it is a tautology. If it is, then ψ *is* a tautological consequence of $\phi_1, \dots, \phi_k$. If it is not, then ψ *is not* a tautological consequence of $\phi_1, \dots, \phi_k$.

PROBLEM 4 *This is important. You should read the solution even if you do not have time to attempt it yourself.*

Show that if ψ is a tautological consequence of the formulas $\phi_1, \dots, \phi_k$ then it is a logical consequence of these formulas. (Trying the case $k = 1$ first might help. Remember that 'ψ is a logical consequence of $\phi_1, \dots, \phi_k$' means whenever $\phi_1, \dots, \phi_k$ are interpreted as being true, then so is ψ.)

SOLUTION 4 We first do the case $k = 1$.

Suppose, then, that ψ is a tautological consequence of ϕ_1. Thus $(\phi_1 \rightarrow \psi)$ is a tautology.

We must show that ψ is a logical consequence of ϕ_1. So suppose we have an interpretation of the non-logical symbols which makes ϕ_1 true. We must show that this interpretation makes ψ true. However, if ψ were false (under this interpretation), $(\phi_1 \rightarrow \psi)$ would also be false by the truth table for implication (table in (d) above) and this contradicts the fact that $(\phi_1 \rightarrow \psi)$ is a tautology (since clearly a tautology is true under all interpretations). This contradiction shows that ψ cannot be false, i.e. it is true, as required.

The general case requires only the following addition. Suppose that we have an interpretation which makes all the formulas $\phi_1, \ldots, \phi_k$ true. Then, by the truth table for conjunction it must make $(\phi_1 \& \phi_2)$ true. Hence it must make $((\phi_1 \& \phi_2) \& \phi_3)$ true. Continuing, we see that is must also make the formula $(\ldots(\phi_1 \& \phi_2) \& \ldots \& \phi_k)$ true. The proof is now similar to the above, using $(\ldots(\phi_1 \& \phi_2) \& \ldots \& \phi_k)$ in place of ϕ_1. ■

5.4 The Tautology Rule and Conditional Proof

Recall that we want the rules of deduction of our formal systems (or logical rules, as we shall usually call them) to be comprehensive enough so that we can formalize as many informal arguments as possible. However, there are two restrictions on the rules we want to introduce, and we discuss this point now.

Firstly, we are obviously trying to formalize *correct* informal arguments, deductions that can lead us only from true statements to true statements. Now this will be guaranteed, according to our discussion in Section 5.3, provided we ensure that our rules of deduction are logically valid. In other words, when a rule of deduction tells us that we may introduce a certain formula into a formal proof depending on certain previously introduced formulas and assumptions, this formula must be a logical consequence of those assumptions and previous formulas.

Notice that this condition is trivially satisfied by the Rule of Assumptions introduced in Section 5.2. For that rule allows us to introduce a formula, ϕ say, on a line in a proof only provided we use ϕ itself as the assumption on that line, and ϕ is clearly a logical consequence of ϕ. (If you look again at the definition of logical consequence in Section 5.3, we hope you will accept this latter statement, even though we have not yet explained an interpretation of the non-logical symbols.)

Our second restriction on formal proofs comes from the observation that any informal proof of a theorem offered by a mathematician must be capable of being checked by (in principle) anyone else. Of course, this is not to say that there need be a general procedure for *producing* the proof—you may recall from the first cassette tape that whether or not a procedure exists for producing proofs of mathematical statements is essentially Leibniz's question, and we shall return to this later—but merely one that will routinely *check* whether or not a purported proof is, in fact, a proof.

We incorporate this property into our formal system by designing each of our logical rules in such a way that it is obvious that there is an algorithm (i.e. a Turing machine) that can check whether or not a given formula can be immediately deduced from other given formulas using that rule. It will thus follow that there is a Turing machine that can check whether or not a given sequence of formulas constitutes a formal proof.

We now specify the second logical rule of our formal system and then say why it satisfies the two restrictions stated above, before giving some examples of its use.

RULE *The Tautology Rule* (*Taut*)
If the formulas $\phi_1, \ldots, \phi_k$ occur in a formal proof and the formula ψ is a tautological consequence of $\phi_1, \ldots, \phi_k$, then on any line subsequent to the lines on which $\phi_1, \ldots, \phi_k$ appear we may introduce the formula ψ. The assumptions on which ψ depends are defined to be all the assumptions on which $\phi_1, \ldots, \phi_k$ depend, and this must be indicated on the left of the line on which ψ is introduced. ▲

By Problem 4 of Section 5.3, the condition that ψ is a tautological consequence of $\phi_1,\ldots,\phi_k$ guarantees that ψ is a logical consequence of $\phi_1,\ldots,\phi_k$, and hence the Tautology Rule is logically valid. Also, the rule specifies that ψ may be immediately deduced from $\phi_1,\ldots,\phi_k$ if and only if ψ is a tautological consequence of $\phi_1,\ldots,\phi_k$ and we know that there is a Turing machine that can check this latter condition. So both our restrictions are satisfied by this rule.

EXAMPLE 1 Here is an example of a formal proof in which the Tautology Rule is used:

1	(1)	$(\exists x_0\ (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}')$	Ass
2	(2)	$(\exists x_0\ (x_0 + x_1) = \mathbf{0} \vee \forall x_3\ x_1 = x_3)$	Ass
1, 2	(3)	$(x_1 = \mathbf{0}' \vee \forall x_3\ x_1 = x_3)$	Taut, (1), (2)

The annotation on the right of line (3) indicates that we have used the Tautology Rule to derive the formula on line (3) from the formulas on lines (1) and (2). Thus the assumptions in force on line (3) are all those in force on lines (1) and (2), namely the formulas numbered 1 and 2. This is indicated by the annotation on the left of the line number. That this is a correct use of the Tautology Rule depends on the fact that the formula

$$(((\exists x_0\ (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}') \,\&\, (\exists x_0\ (x_0 + x_1) = \mathbf{0} \vee \forall x_3\ x_1 = x_3)) \\ \rightarrow (x_1 = \mathbf{0}' \vee \forall x_3\ x_1 = x_3))$$

is a tautology, and this we have already checked in Problem 2 of Section 5.3.

The above table thus constitutes a formal proof of the formula on line (3), namely $(x_1 = \mathbf{0}' \vee \forall x_3\ x_1 = x_3)$, depending on the assumptions in force on line (3), namely the formulas on lines (1) and (2), i.e. $(\exists x_0\ (x_0 + x_1) = \mathbf{0} \rightarrow x_1 = \mathbf{0}')$ and $(\exists x_0\ (x_0 + x_1) = \mathbf{0} \vee \forall x_3\ x_1 = x_3)$. ☐

We mention in passing that an informal version of the Tautology Rule was used in the informal proof (third version) given in Section 5.2. For consider the step made at line (11). Suppose the formula ϕ_1 corresponds to the informal assertion 'n^2 is even' and ϕ_2 corresponds to 'n^2 is odd'—we have not yet, of course, explained exactly what this means, but this does not matter for the moment: it is merely the form of the deduction that we are illustrating. Since it is the symbol & of our formal language that corresponds to the English word 'and', it is the formula $(\phi_1 \,\&\, \phi_2)$ that will correspond to the informal assertion 'n^2 is even and n^2 is odd'. Now $((\phi_1 \,\&\, \phi_2) \rightarrow (\phi_1 \,\&\, \phi_2))$ is a tautology (as you can easily check), so we have indeed applied the Tautology Rule at line (11) (with $k = 2$ and $\psi = (\phi_1 \,\&\, \phi_2)$).

Our next logical rule also corresponds to an informal one used in the informal proof of Section 5.2, namely the one used at line (14), where we incorporated an assumption into the assertion by use of the word 'implies'. Since it is the symbol → that corresponds to the word 'implies', we obtain our third rule, which is called *Conditional Proof*, as follows.

RULE *Conditional Proof* (*CP*)

If a formula, ψ say, occurs on a line in a formal proof, and a formula, ϕ say, occurs amongst the assumptions in force on that line, then on any subsequent line we may introduce the formula $(\phi \rightarrow \psi)$. The assumptions on which $(\phi \rightarrow \psi)$ depends, which must be written to the left of $(\phi \rightarrow \psi)$, are all those on which ψ depends except for ϕ. ▲

To see that this rule is logically valid, we must show that if ψ is a logical consequence of $\phi, \phi_1,\ldots,\phi_k$ (where these are all the assumptions appearing to the left of ψ), then $(\phi \rightarrow \psi)$ is a logical consequence of $\phi_1,\ldots,\phi_k$. This will clearly demonstrate that the rule CP can lead only from logically valid lines in a proof to logically valid lines. So suppose that the formulas $\phi_1, \ldots, \phi_k$ are all true under some interpretation of the non-logical symbols—again, it does not matter exactly what this means since it is only truth tables for the connectives that are needed here. We must show that $(\phi \to \psi)$ is true. Suppose it is false. Then by the truth table for →, we must have ϕ true and ψ false. (This is the only way $(\phi \to \psi)$ can be false.) But now $\phi, \phi_1, \ldots, \phi_k$ are all true and ψ is false, which contradicts the fact that ψ is a logical consequence of $\phi, \phi_1, \ldots, \phi_k$. Thus $(\phi \to \psi)$ is true, as required.

CP also meets our second restriction on logical rules, since to check whether it has been used correctly on a given line in a given proof, a machine has to check merely whether the formula on that line is of the form $(\phi \to \psi)$ and take note of the assumptions on that line—say they are $\phi_1, \ldots, \phi_k$. The machine then checks all previous lines in the proof until it finds the formula ψ with assumptions $\phi_1, \ldots, \phi_k, \phi$. If it does not find such a line, CP has not been applied correctly; otherwise it has.

XAMPLE 2 We give a formal proof of the formula

$$(x_0 = x_1 \to (x_0 = x_1 \lor x_1 = \mathbf{0}))$$

depending on no assumptions at all.

1	(1)	$x_0 = x_1$	Ass
1	(2)	$(x_0 = x_1 \lor x_1 = \mathbf{0})$	Taut, (1)
	(3)	$(x_0 = x_1 \to (x_0 = x_1 \lor x_1 = \mathbf{0}))$	CP, (2)

The use of the Tautology Rule on line (2) is justified since $(x_0 = x_1 \to (x_0 = x_1 \lor x_1 = \mathbf{0}))$ is a tautology, as you can easily check by calling the subformulas $x_0 = x_1$ and $x_1 = \mathbf{0}$, θ and χ respectively, so that the formula becomes $(\theta \to (\theta \lor \chi))$. The use of CP on line (3) is justified as follows. On line (2) the formula $(x_0 = x_1 \lor x_1 = \mathbf{0})$ appears with the assumption $x_0 = x_1$. CP tells us that we may derive $(x_0 = x_1 \to (x_0 = x_1 \lor x_1 = \mathbf{0}))$, which will depend on those assumptions used to derive $(x_0 = x_1 \lor x_1 = \mathbf{0})$, other than $x_0 = x_1$ itself. But there are no such assumptions. Hence on line (3) we write nothing to the left of the line number, and CP, (2) to the right of the formula to show that CP has been used to deduce the formula on line (3) from line (2).

XAMPLE 3 In this example we show that if ϕ, ψ are any formulas, then there is a formal proof of the formula $(-\psi \to \phi)$ depending on the assumption ψ.

1	(1)	ψ	Ass
2	(2)	$-\psi$	Ass
1, 2	(3)	ϕ	Taut, (1), (2)
1	(4)	$(-\psi \to \phi)$	CP, (3)

The use of the Tautology Rule on line (3) is justified since $((\psi \mathbin{\&} -\psi) \to \phi)$ is a tautology (as you can easily check) and, as dictated by the rule, the assumptions on which ϕ depends are all those on which ψ, $-\psi$ depend, and we have indicated these to the left of line (3). Now the assumptions on line (3) include the formula $-\psi$ (since the number 2 appears on the left), so CP allows us to introduce the formula $(-\psi \to \phi)$ on any subsequent line, provided we maintain all the assumptions of line (3) except for $-\psi$, and this we have done on line (4).

XAMPLE 4 Here is a formal proof in which the rule CP is applied twice. As in Example 3, ϕ, ψ are any formulas, and we give a formal proof of the formula $(\phi \to (\psi \to (\phi \mathbin{\&} \psi)))$ depending on no assumptions.

1	(1)	ϕ	Ass
2	(2)	ψ	Ass
1, 2	(3)	$(\phi \mathbin{\&} \psi)$	Taut, (1), (2)
1	(4)	$(\psi \to (\phi \mathbin{\&} \psi))$	CP, (3)
	(5)	$(\phi \to (\psi \to (\phi \mathbin{\&} \psi)))$	CP, (4)

We have used on line (3) the fact that $((\phi \,\&\, \psi) \rightarrow (\phi \,\&\, \psi))$ is a tautology. On line (4) we have used CP to transfer the formula ψ from the assumptions of line (3), and on line (5) we have used CP to transfer the formula ϕ from the assumptions of line (4). Thus we have given a formal proof of the formula $(\phi \rightarrow (\psi \rightarrow (\phi \,\&\, \psi)))$ depending on no assumptions. ☐

The examples above illustrate an important point about the strategy used for finding formal proofs. When you are required to prove a formula of the form $(\phi \rightarrow \psi)$, add the formula ϕ to your list of assumptions and then prove ψ. Then the rule CP permits you to derive $(\phi \rightarrow \psi)$ from your original assumptions. You will see more examples of this strategy in the next unit.

In Example 4 we effectively used this strategy twice. To prove $(\phi \rightarrow (\psi \rightarrow (\phi \,\&\, \psi)))$ from no assumptions, we first added ϕ to our list of assumptions and set out to prove $(\psi \rightarrow (\phi \,\&\, \psi))$. But then we used the strategy a second time, adding ψ to our list of assumptions (to get the list ϕ, ψ), and tried to prove $(\phi \,\&\, \psi)$ from these assumptions. We duly completed this proof, and CP used twice gave the proof of the required formula.

PROBLEM 1 The following are incomplete formal proofs because the assumption numbers are missing. Fill in these missing numbers and where the tautology rule has been applied, state which tautologies have been used. (ϕ, ψ, χ are any formulas.)

(i)	1	(1)	ψ	Ass
	2	(2)	ϕ	Ass
	1	(3)	$(\phi \vee \psi)$	Taut, (1)
	1, 2	(4)	$(\phi \,\&\, \psi)$	Taut, (1), (2)
	1, 2	(5)	$((\phi \vee \psi) \,\&\, (\phi \,\&\, \psi))$	Taut, (3), (4)
	2	(6)	$(\psi \rightarrow ((\phi \vee \psi) \,\&\, (\phi \,\&\, \psi)))$	CP, (5)
(ii)		(1)	$(\phi \vee \chi)$	Ass
		(2)	$-\chi$	Ass
		(3)	ϕ	Taut, (1), (2)
		(4)	ψ	Ass
		(5)	$(\phi \,\&\, \psi)$	Taut, (3), (4)
		(6)	$(-\chi \rightarrow (\phi \,\&\, \psi))$	CP, (5)
		(7)	$(\chi \vee (\phi \,\&\, \psi))$	Taut, (6)
		(8)	$(\psi \rightarrow (\chi \vee (\phi \,\&\, \psi)))$	CP, (7)

PROBLEM 2 (i) Write down a formal proof of the formula $-(\phi \rightarrow \psi)$ depending on the assumptions ϕ, $-\psi$.

(ii) How would you turn the proof produced for part (i) into a formal proof of the formula $(\phi \rightarrow (-\psi \rightarrow -(\phi \rightarrow \psi)))$ depending on no assumptions?

PROBLEM 3 Write down a formal proof of the formula $(\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi))$ depending on no assumptions, using the fact that $((\phi \,\&\, (\phi \rightarrow \psi)) \rightarrow \psi)$ is a tautology.

SOLUTION 1 (i) The assumption numbers are:

1 on line (1);

2 on line (2);

1 on line (3), where the tautology $(\psi \rightarrow (\phi \vee \psi))$ has been used;

1, 2 on line (4), where the tautology $((\phi \,\&\, \psi) \rightarrow (\phi \,\&\, \psi))$ has been used;

1, 2 on line (5), where the tautology $(((\phi \vee \psi) \,\&\, (\phi \,\&\, \psi)) \rightarrow ((\phi \vee \psi) \,\&\, (\phi \,\&\, \psi)))$ has been used;

2 on line (6).

(ii) The assumption numbers are:

1 on line (1);

2 on line (2);

1, 2 on line (3), where the tautology $(((\phi \vee \chi) \,\&\, -\chi) \rightarrow \phi)$ has been used;

4 on line (4);

1, 2, 4 on line (5), where the tautology $((\phi \,\&\, \psi) \rightarrow (\phi \,\&\, \psi))$ has been used;

1, 4 on line (6);

1, 4 on line (7), where the tautology $((-\chi \rightarrow (\phi \,\&\, \psi)) \rightarrow (\chi \vee (\phi \,\&\, \psi)))$ has been used;

1 on line (8). ■

LUTION 2 (i)

1	(1)	ϕ	Ass
2	(2)	$-\psi$	Ass
1, 2	(3)	$-(\phi \rightarrow \psi)$	Taut, (1), (2)

Tautology used in line (3): $((\phi \,\&\, -\psi) \rightarrow -(\phi \rightarrow \psi))$. There are other possible correct proofs.

(ii) All that is needed is two uses of CP. For the proof given in the solution to part (i) above, we just add the lines:

1	(4)	$(-\psi \rightarrow -(\phi \rightarrow \psi))$	CP, (3)
	(5)	$(\phi \rightarrow (-\psi \rightarrow -(\phi \rightarrow \psi)))$	CP, (4) ■

LUTION 3

1	(1)	ϕ	Ass
2	(2)	$(\phi \rightarrow \psi)$	Ass
1, 2	(3)	ψ	Taut, (1), (2)
1	(4)	$((\phi \rightarrow \psi) \rightarrow \psi)$	CP, (3)
	(5)	$(\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi))$	CP, (4) ■

This concludes the formal rules governing the connectives. In the next unit we shall study the variables and quantifiers in more detail and explain the rules governing them. We shall then be able to produce much more interesting formal proofs.

5.5 Objectives of Unit 5

We list those things on which we may set assessment questions to test your understanding of the unit. After working through the unit you should be able to:

(i) recall the definitions of a term, an atomic formula and a formula of our formal language;

(ii) determine whether a given sequence of symbols is a term, an atomic formula, or a formula of our formal language;

(iii) find all the subformulas of a given formula;

(iv) write out a truth table for a given formula, and determine whether a given formula is a tautological consequence of other given formulas;

(v) recall what each of the rules Ass, Taut, and CP says;

(vi) check a purported formal proof to see whether or not the uses of these rules in it are legitimate;

(vii) construct simple formal proofs using these rules.

5.6 Appendix: A Turing Machine for Terms

The following material is optional and non-assessed.

We wish to devise a Turing machine that will recognize whether or not a given string of our basic symbols (i.e. those symbols listed in Section 5.1) is a term of our formal language. As we have mentioned already, it will be important for later work to know that such a machine exists, although the exact details of its machine table will be of no consequence. Hence we merely give an outline of how such a machine could work.

As a Turing machine can cope with only a finite list of symbols, we shall adopt the convention, mentioned in Section 5.1, of representing the variable x_n by $x_{\underbrace{\prime\prime\prime\ldots\prime\prime}_{n}}$ i.e. x with a subscript composed of n primes.

The idea is to mimic the analysis of terms described above by having our machine break down a given string to see whether it was built up from variables and the zero symbol using clause (2) in the definition of a term. To help us, we shall need the following definition.

DEFINITION If τ is a term other than a variable or the zero symbol, the *leading symbol occurrence in* τ (l.s.o. in τ, for short) is defined to be the last occurrence of either $+$, $\cdot$ or $'$ used in building up τ using clause (2) of the definition of a term. In other words, the l.s.o. in τ is the occurrence of $+$, $\cdot$ or $'$ that would be analysed first in the term analysis previously described. ▲

Thus, for example, the l.s.o. in $(\mathbf{0} + (x_{\prime} + x_{\prime\prime}))$ is the left-most occurrence of the symbol $+$, while the l.s.o. in $(\mathbf{0} + x_{\prime})'$ is the (only) occurrence of the symbol $'$.

How could a Turing machine find the leading symbol occurrence in a term that is written on its tape? If the term ends with the successor symbol, then clearly this must be the l.s.o. in the term and it is a simple matter to programme a Turing machine to detect this. The difficulty occurs when the l.s.o. is an occurrence of $+$ or $\cdot$ buried somewhere in the middle of the term. In this case, we base our algorithm on the following observation. Firstly, given any occurrence of the symbol $+$ or $\cdot$ in a term τ, let L denote the number of left-hand brackets, i.e. (s, occurring to the left of this occurrence and let R denote the number of right-hand brackets, i.e.)s, occurring to the left of this occurrence. For example, for the first occurrence of the symbol $+$ in the term $((\mathbf{0} + x_{\prime}) + x_{\prime\prime})$ we have $L = 2$, $R = 0$, and for the second occurrence, $L = 2$, $R = 1$. Then one can show that an occurrence of $+$ or $\cdot$ in a term is the l.s.o. of that term if and only if $R + 1 = L$. We do not bother to give a proof of this fact, although looking at some examples should make it seem plausible. Thus consider the term $(((\mathbf{0} \cdot x_{\prime}) \cdot x_{\prime})'' + (x_{\prime\prime} + \mathbf{0}''))$. This term contains two occurrences of the symbol $\cdot$ followed by two occurrences of the symbol $+$. The corresponding numbers L, R are (from left to right): $L = 3$, $R = 0$; $L = 3$, $R = 1$; $L = 3$, $R = 2$; $L = 4$, $R = 2$. The only pair here which satisfies $R + 1 = L$ is $L = 3$, $R = 2$, so the result we have just stated predicts that the first occurrence of the symbol $+$ must be the l.s.o. in this term, and this is easily confirmed by the usual analysis.

Now clearly a Turing machine can be devised to compute the numbers L, R for each occurrence of the symbols $+$, $\cdot$ in a string of symbols written on its tape, and to check whether exactly one of these pairs satisfies $R + 1 = L$. Hence we have a machine that can locate the l.s.o., as required.

It is now an easy matter to devise a Turing machine that can recognize whether or not a string of symbols written on its tape is a term. ('Recognize' here means the machine will halt with, say, a single 1 on its tape if the string is a term and halt with a blank tape otherwise.) Such a machine first looks for the l.s.o. using the algorithm just described. If this is the symbol $'$ (located at the end of the string), then the machine deletes this symbol and repeats the procedure with the remaining string. If the l.s.o. is the symbol $+$ or $\cdot$ then the machine checks that the string begins with a left bracket, which it then deletes, and ends with a right bracket, which it also deletes, and repeats this entire procedure on both the string to the left and the string to the right of the l.s.o. When a string with no occurrences of symbols $+$, $\cdot$ or $'$ is reached, the machine checks that this is either the zero symbol or a variable, i.e. a string of the form $x_{,,\ldots,}$. If, at any stage, this procedure breaks down (e.g. because the machine finds a string with no, or two or more, occurrences of $+$ or $\cdot$ which satisfy $R + 1 = L$), then the original string could not have been a term. Otherwise, the original string was a term.

Let us apply this algorithm to the string $((x_{,} \cdot x_{,}) + \mathbf{0})'$. The l.s.o. here is the symbol $'$, so this is deleted to obtain $((x_{,} \cdot x_{,}) + \mathbf{0})$. Since this string does not end with the symbol $'$ we apply the 'counting brackets' argument to each occurrence of $\cdot$ and $+$. For the only occurrence of $\cdot$ we have $L = 2$, $R = 0$, and for the only occurrence of $+$ we have $L = 2$, $R = 1$. Hence there is indeed exactly one occurrence of $+$ or $\cdot$ such that $R + 1 = L$, namely, the only occurrence of $+$. Thus we check that the string begins with a (and ends with a), delete these brackets, and then repeat the procedure on the strings $(x_{,} \cdot x_{,})$ and $\mathbf{0}$. Now $\mathbf{0}$ is the zero symbol, so it remains only to apply the procedure to $(x_{,} \cdot x_{,})$. The occurrence of $\cdot$ here certainly satisfies $R + 1 = L$ ($L = 1$, $R = 0$), so we again check the outermost brackets, delete them, and reapply the procedure to each of the strings $x_{,,}$, $x_{,}$. But these are variables, so the algorithm is completed and the original string was indeed a term.

Consider now the string $((x_{,} \cdot x_{,,} + x_{,}) + \mathbf{0})$. Since this does not end with the symbol $'$, we 'count brackets' to obtain, for the occurrences of $+$ and $\cdot$ (from left to right): $L = 2$, $R = 0$; $L = 2$, $R = 0$; $L = 2$, $R = 1$. Thus the l.s.o. is the second occurrence of $+$ and, after checking and deleting the outermost brackets, we must repeat the procedure on the strings $(x_{,} \cdot x_{,,} + x_{,})$ and $\mathbf{0}$. However, 'counting brackets' for occurrences of $+$ and $\cdot$ in the first string here yields $L = 1$, $R = 0$ for both occurrences. Hence there is not a unique occurrence of $+$ or $\cdot$ satisfying $R + 1 = L$, so the algorithm halts at this point having revealed that the original string was not a term.

As we have said already, we hope it is clear that we have given an algorithm for checking terms which could be used on a Turing machine, and we are not giving the machine table of such a machine.

UNIT 6
QUANTIFIER LOGIC

6.0 Introduction

In *Unit 5* we described the language of number theory and introduced some of the logical rules of the corresponding formal system. We emphasized two key features of these rules:

(a) there is an algorithm (Turing machine) to decide whether a given sequence of formulas is a formal proof from a given set of assumptions; and

(b) the rules are logically valid; that is, if a formula, ψ say, occurs on a line in a formal proof depending on assumptions $\phi_1, \ldots, \phi_k$, say, then any interpretation of the non-logical symbols which makes each of the formulas $\phi_1, \ldots, \phi_k$ true also makes ψ true.

We have not completely explained (b) since we have not yet told you what an interpretation of the non-logical symbols looks like, but we did explain the interpretation of the connectives (using truth tables) which was sufficient to establish the logical validity of the Tautology Rule and Conditional Proof. The Rule of Assumptions is trivially logically valid.

In this unit we complete the description of our formal system by giving the logical rules for the quantifiers $\forall$ and $\exists$, and for the identity symbol $=$.

The intended interpretation of these symbols should be clear. There are just two points worth noting. Firstly, although $\forall x$ is interpreted as 'for every x', this does not mean 'every x that exists' but only 'every x in the set being considered'; and when we give the interpretation we need to specify what this set is. So the formula $\forall x\ \forall y\ (x \cdot y) = (y \cdot x)$ can be interpreted as expressing the fact that multiplication of real numbers is commutative if we take the universal quantifiers as referring to the set of real numbers. But the same formula can also be interpreted as expressing the fact that multiplication of integers is commutative if we take the universal quantifiers as referring to the set of integers. The traditional terminology for the set of objects under consideration is the *universe of discourse*. Another term for this set is the *domain* of the interpretation and we shall follow **BJ** in using this expression.

Thus when we want to give an interpretation of our symbols we need to specify which set is the domain of the interpretation. Then the existential quantifier $\exists x$ is interpreted as saying 'there is some x in the domain'.

Secondly, the identity symbol in the formula $\tau_1 = \tau_2$ is interpreted as expressing the fact that the objects referred by τ_1 and τ_2 are equal. So, for example, with the standard arithmetical interpretation, the formula $(\mathbf{0}'' \cdot \mathbf{0}''') = \mathbf{0}''''''$ is interpreted as being true. The terms in this formula are different expressions, but with the usual interpretation of the arithmetical symbols they represent the same number, namely the number 6.

We shall look further at the interpretation of the symbols in *Unit 7*.

Just as the formulation of the Tautology Rule required an analysis of the subformulas of a given formula, so we shall have to investigate more deeply another aspect of the structure of formulas before we can formulate the quantifier rules. Namely, we shall begin by looking at the way variables occur in formulas.

6.1 Free and Bound Variables

In everyday mathematics variables are used in several different ways. Let us illustrate this with the following situation. Suppose we are given a set S, with two binary operations $+$ and $\cdot$ defined on S. (Thus S might be the set of natural numbers, and the operations those of ordinary addition and multiplication, or S might be the set of real numbers, or a set of matrices etc.) Now, there are many mathematical statements we could make about such a set-up, for example, 'the operation $+$ is commutative', 'the operation $\cdot$ is commutative', 'the operations $+$ and $\cdot$ are distributive'. We can express these statements more formally using variables and the formal symbols $+$ and $\cdot$, as, respectively:

$$\forall x\ \forall y\ (x + y) = (y + x) \tag{1}$$

$$\forall x\ \forall y\ (x \cdot y) = (y \cdot x) \tag{2}$$

$$\forall x\ \forall y\ \forall z\ (x \cdot (y + z)) = ((x \cdot y) + (x \cdot z)) \tag{3}$$

Consider also the statement:

$$((x \cdot x) + (y \cdot y)) = 1 \tag{4}$$

(where we are here supposing that S is, say, the set of real numbers).

Do you see an essential difference in the use of the variables in (1), (2) and (3) from that in (4)? In (1), for example, the statement $\forall x\ \forall y\ (x + y) = (y + x)$ is a statement about the operation $+$ (namely it is asserting that $+$ is commutative) and not about x or y, whereas the statement (4) is telling us something about x and y, namely, that the point (x, y) lies on the circle with centre $(0, 0)$ and radius 1. The variables in (1), (2) and (3) are often called *dummy* variables. We shall also say that the occurrences of the variables in (1), (2) and (3) are *bound* occurrences, and the occurrences of the variables in (4) are *free* occurrences. (We shall see later that they are called free occurrences because we are free to make substitutions for these occurrences. In contrast we use 'bound' for the other occurrences.)

Another way of looking at the uses of the variables in (1), (2), (3) and (4) which might help in distinguishing them is as follows. Ask yourself the question 'is the formula true?' and note the information required for an answer. Thus, to answer 'is $\forall x\ \forall y\ (x \cdot y) = (y \cdot x)$ true?', you need only know what kind of multiplication is intended to interpret the symbol $\cdot$. If multiplication of real numbers is intended (as in our original example), the answer is 'yes'. If multiplication of say, 2×2 matrices with real number entries is intended, the answer is 'no'. However, in order to answer the question 'is $((x \cdot x) + (y \cdot y)) = 1$ true?', you need not only information about which operations $+$ and $\cdot$ are intended to be (and on which sets they are operating) but also you need to know what x and y are. This extra information required (i.e. the response 'it depends on what x and y are') distinguishes the occurrences of x and y in (4) as free occurrences from those in (2), where they are bound occurrences.

The same formula can contain both free and bound occurrences of variables. For example, in

$$\exists y\ x = (y + y) \tag{5}$$

the occurrence of x is free, but the occurrences of y are all bound; the question 'is $\exists y\ x = (y + y)$ true?' produces the response 'it depends on what x is' but not 'it depends on what y is'. For example, if we are told that the $+$ symbol here is referring to addition of positive integers, then the formula $\exists y\ x = (y + y)$ is true if $x = 6$ but false if $x = 17$. In other words, the formula (5) tells us something about x, but not about y. We could equally well express the same fact about x by

$$\exists t\ x = (t + t)$$

in which the variable y does not occur.

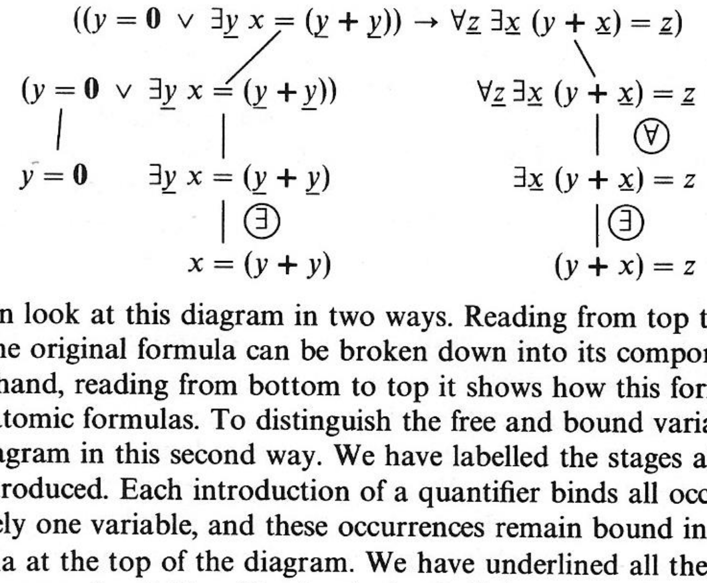
It is, of course, the presence of the quantifiers that converts free occurrences of variables into bound occurrences. Thus if we trace how a formula is built up from atomic formulas we can work out which variables are free and which are bound. Before we give an example, however, let us introduce the following convention.

CONVENTION A variable has been defined as one of the symbols $x_0, x_1, x_2, \ldots$. This precision was needed in order that a Turing machine could be constructed to recognize basic symbols and decide which sequences of them were terms, formulas etc. Now that this has been accomplished we shall err in favour of typographical convenience and clarity when writing down formulas. Thus we shall use $x, y, z, t, \ldots$ for variables. If you like, you can regard $x, y, z, t, \ldots$ as abbreviations for $x_0, x_1, x_2, x_3, \ldots$. All that really matters is that they represent distinct variables.

To return to the matter at hand we consider the usual analysis of the formula

$$((y = \mathbf{0} \lor \exists y\ x = (y + y)) \rightarrow \forall z\ \exists x\ (y + x) = z)$$

It is as follows:



$((y = \mathbf{0} \lor \exists \underline{y}\ x = (\underline{y} + \underline{y})) \rightarrow \forall \underline{z}\ \exists \underline{x}\ (y + \underline{x}) = \underline{z})$
$(y = \mathbf{0} \lor \exists \underline{y}\ x = (\underline{y} + \underline{y}))$
$\forall \underline{z}\ \exists \underline{x}\ (y + \underline{x}) = \underline{z}$
$y = \mathbf{0}$
$\exists \underline{y}\ x = (\underline{y} + \underline{y})$
$\exists \underline{x}\ (y + \underline{x}) = z$
ⓔ
Ⓐ
$x = (y + y)$
$(y + x) = z$

We can look at this diagram in two ways. Reading from top to bottom it shows how the original formula can be broken down into its components. On the other hand, reading from bottom to top it shows how this formula is built up from atomic formulas. To distinguish the free and bound variables we look at the diagram in this second way. We have labelled the stages at which quantifiers are introduced. Each introduction of a quantifier binds all occurrences of precisely one variable, and these occurrences remain bound in the original formula at the top of the diagram. We have underlined all the bound occurrences of variables. Clearly, the 'underlining process' we have carried out here is purely mechanical, and hence it provides us with an algorithm for deciding whether or not a given occurrence of a variable in a formula is free or bound. Notice that we have been careful to say that *occurrences* of variables are free or bound, for, as this example shows, the same variable can *occur* free and bound (in different places) in the same formula. For consider the variable y in the formula we have just dealt with:

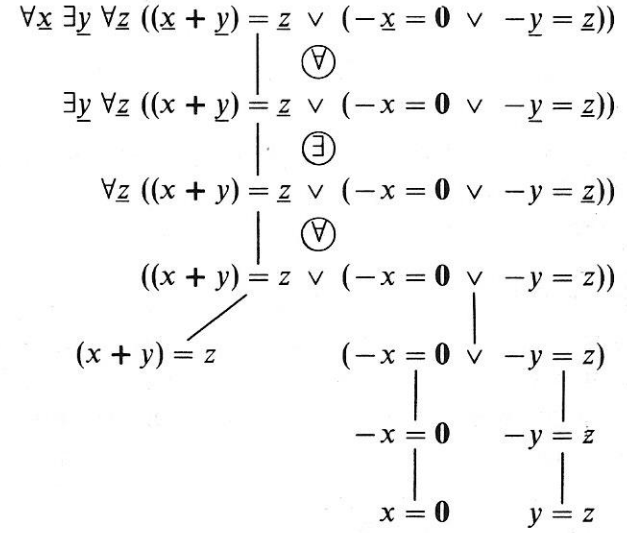
$$((\underset{\overset{\uparrow}{1}}{y} = \mathbf{0} \lor \exists \underset{\overset{\uparrow}{2}}{y}\ x = (\underset{\overset{\uparrow}{3}}{y} + \underset{\overset{\uparrow}{4}}{y})) \rightarrow \forall z\ \exists x\ (\underset{\overset{\uparrow}{5}}{y} + x) = z).$$

Its first and fifth occurrences are free, whereas its second, third and fourth occurrences are bound.

After some practice, it should be possible for you to distinguish between free and bound occurrences of variables without the need to actually write down the analysis of the formula. In the next section we shall see another use of this analysis.

XAMPLE 1 $\forall x\,\exists y\,\forall z\,((x + y) = z \vee (-x = \mathbf{0} \vee -y = z))$

The analysis is as follows:


$\forall\underline{x}\,\exists\underline{y}\,\forall\underline{z}\,((\underline{x} + \underline{y}) = \underline{z} \vee (-\underline{x} = \mathbf{0} \vee -\underline{y} = \underline{z}))$
Ⓥ
$\exists\underline{y}\,\forall\underline{z}\,((x + \underline{y}) = \underline{z} \vee (-x = \mathbf{0} \vee -\underline{y} = \underline{z}))$
∃
$\forall\underline{z}\,((x + y) = \underline{z} \vee (-x = \mathbf{0} \vee -y = \underline{z}))$
Ⓥ
$((x + y) = z \vee (-x = \mathbf{0} \vee -y = z))$
$(x + y) = z$
$(-x = \mathbf{0} \vee -y = z)$
$-x = \mathbf{0}$
$-y = z$
$x = \mathbf{0}$
$y = z$

Thus all occurrences of x, y and z in the above formula are bound occurrences. ☐

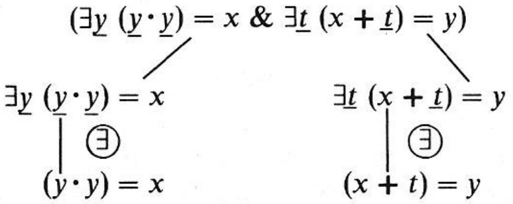
(Remember that $-$ is the negation symbol and not a 'minus' symbol, so that the formula $-y = z$ should be read as 'it is not the case that y equals z', rather than 'minus y equals z'.)

XAMPLE 2 (i) $(\exists y\,(y \cdot y) = x \,\&\, \exists t\,(x + t) = y)$

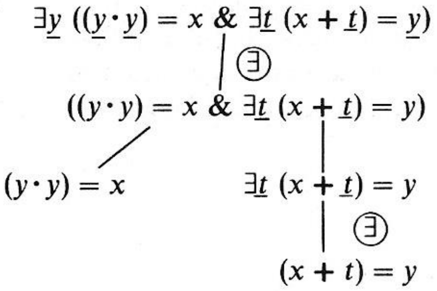
(ii) $\exists y\,((y \cdot y) = x \,\&\, \exists t\,(x + t) = y)$

Here are the analyses:

(i)


$(\exists\underline{y}\,(\underline{y} \cdot \underline{y}) = x \,\&\, \exists\underline{t}\,(x + \underline{t}) = y)$
$\exists\underline{y}\,(\underline{y} \cdot \underline{y}) = x$
$\exists\underline{t}\,(x + \underline{t}) = y$
∃
∃
$(y \cdot y) = x$
$(x + t) = y$

(ii)


$\exists\underline{y}\,((\underline{y} \cdot \underline{y}) = x \,\&\, \exists\underline{t}\,(x + \underline{t}) = \underline{y})$
∃
$((y \cdot y) = x \,\&\, \exists\underline{t}\,(x + \underline{t}) = y)$
$(y \cdot y) = x$
$\exists\underline{t}\,(x + \underline{t}) = y$
∃
$(x + t) = y$

These examples show how important bracketing can be. In (i), the first, second and third occurrences of y are bound, whereas the fourth occurrence is free. In (ii), all four occurrences of y are bound. In (i) and (ii) both occurrences of x are free and both occurrences of t are bound. ☐

Notice that in Example 1 there are no free occurrences of any variable. Formulas with this property are called *sentences*. Thus a sentence is a formula in which all occurrences of variables are bound.

ROBLEM 1 In the following formulas, determine which occurrences of variables are free and which are bound. State which of the formulas are sentences.

(i) $\forall x\,\exists y\; x = y$

(ii) $((x = y \vee \exists t\,(x + t) = y) \vee \exists t\,(y + t) = x)$

(iii) $(\exists z\ (z \cdot x) = y \to \forall y\ (\exists z\ (z \cdot y) = x \to -\exists x\ (x \cdot x) = y))$

(iv) $\exists x\ (x = y \vee x = y)$

(v) $(\exists x\ x = y \vee x = y)$

SOLUTION 1 (i) Both occurrences of x and y are bound. Since no other variables occur, the formula is a sentence.

(ii) All occurrences of x and y are free, and all occurrences of t are bound. The formula is not a sentence.

(iii)
$$(\exists \underset{\overset{\uparrow}{\mathrm{B}}}{z}\ (\underset{\overset{\uparrow}{\mathrm{B}}}{z}\cdot \underset{\overset{\uparrow}{\mathrm{F}}}{x}) = \underset{\overset{\uparrow}{\mathrm{F}}}{y} \to \forall \underset{\overset{\uparrow}{\mathrm{B}}}{y}\ (\exists \underset{\overset{\uparrow}{\mathrm{B}}}{z}\ (\underset{\overset{\uparrow}{\mathrm{B}}}{z}\cdot \underset{\overset{\uparrow}{\mathrm{B}}}{y}) = \underset{\overset{\uparrow}{\mathrm{F}}}{x} \to -\exists \underset{\overset{\uparrow}{\mathrm{B}}}{x}\ (\underset{\overset{\uparrow}{\mathrm{B}}}{x}\cdot \underset{\overset{\uparrow}{\mathrm{B}}}{x}) = \underset{\overset{\uparrow}{\mathrm{B}}}{y}))$$

Here F indicates a free occurrence of a variable and B a bound occurrence. The formula is not a sentence.

(iv) Both occurrences of y are free; all three occurrences of x are bound. The formula is not a sentence.

(v) The first and second occurrences of x are bound; both occurrences of y are free. The formula is not a sentence. ■

6.2 The Universal Quantifier Elimination Rule

We are now in a position to begin describing the formal rules for the quantifiers, and we start with the universal quantifier ∀. We have two rules associated with this symbol: one for eliminating it from proofs and the other for introducing it into proofs. We first consider the elimination rule.

Our method for devising the quantifier rules will be to look at how quantifiers are handled in informal mathematical arguments and to mirror this in our formal rules. So we begin by looking at some examples. Consider the following extract:

> 'From the identity $x^2 - 1 = (x - 1)(x + 1)$ it follows that $7^2 - 1 = (7 - 1)(7 + 1)$ and so is divisible by 8'.

At first sight, there is no universal quantifier to be seen, but it must be remembered that in informal mathematics the symbol ∀ is not often used and other methods are used to indicate that statements are universally true. In the example above, the author has used the word 'identity' to indicate that the equation which follows is true for all values (in the range considered) of the variable x. So the deduction in this passage, expressed more formally, is as follows. From

$$\forall x\ x^2 - 1 = (x - 1)\ (x + 1) \tag{1}$$

we deduce that

$$7^2 - 1 = (7 - 1)(7 + 1). \tag{2}$$

We can see that in going from (1) to (2) the universal quantifier ∀ and the following variable x have been eliminated, and all the remaining occurrences of x have been replaced by occurrences of the symbol 7; the variable which expressed generality in (1) has been replaced by a symbol for a particular number. However, we are not obliged to replace the variable by a numeral; it could be replaced by a more complicated expression. For example, in an argument in trigonometry we might want to deduce from (1) that

$$(\sin t)^2 - 1 = (\sin t - 1)(\sin t + 1). \tag{3}$$

So this time x has been replaced by $\sin t$, and it is easy to imagine a situation where we might want to substitute even more complicated expressions for x. In general, the sort of expressions we might want to substitute for x will be those which can stand for numbers. Recall that in *Unit 5* we called these numerical expressions *terms*. So the rule we are looking for will be one which allows us to eliminate a universal quantifier and the following variable from the beginning of a formula and to replace all the remaining now free (but previously bound) occurrences of the variable by occurrences of some term.

TATION In order to be able to state this rule succinctly, it is convenient to introduce some notation. If ϕ is a formula, x a variable and τ a term, then $\phi(\tau/x)$ denotes the result of taking the formula ϕ and replacing all free occurrences of x by occurrences of the term τ.

AMPLE 1

(i) If ϕ is the formula $\exists y\ y = (x + x)$ then $\phi(\mathbf{0}/x)$ is the formula $\exists y\ y = (\mathbf{0} + \mathbf{0})$ and $\phi((x \cdot z)/x)$ is the formula $\exists y\ y = ((x \cdot z) + (x \cdot z))$.

(ii) If ϕ is the formula $(\forall y\ (x \cdot y) = x \rightarrow \forall x\ (x \cdot y) = x)$ then $\phi(\mathbf{0}'/x)$ is the formula

$$(\forall y\ (\mathbf{0}' \cdot y) = \mathbf{0}' \rightarrow \forall x\ (x \cdot y) = x)$$

since we replace only free occurrences of x by $\mathbf{0}'$ and the last three occurrences of x in ϕ are bound occurrences. For a similar reason, $\phi(\mathbf{0}'/y)$ is the formula

$$(\forall y\ (x \cdot y = x) \rightarrow \forall x\ (x \cdot \mathbf{0}') = x).$$

OBLEM 1 In the following cases write down the formula $\phi(\tau/x)$.

(i) ϕ is $\exists y\ (y \cdot y) = (x \cdot (x + y))$; τ is $(z'' + \mathbf{0})$.

(ii) ϕ is $(\exists x\ (x \cdot x) = y \ \&\ x = (z + y))$; τ is $\mathbf{0}'$.

(iii) ϕ is $\exists x\ ((x \cdot x) = y \ \&\ x = (z + y))$; τ is $\mathbf{0}'$.

(iv) ϕ is $\exists y\ y = (x + x)$; τ is y'.

(v) ϕ is $\exists z\ (x + z) = (x \cdot z)$; τ is x.

LUTION 1

(i) $\exists y\ (y \cdot y) = ((z'' + \mathbf{0}) \cdot ((z'' + \mathbf{0}) + y))$

(ii) $(\exists x\ (x \cdot x) = y \ \&\ \mathbf{0}' = (z + y))$

Note that the first three occurrences of x in the formula ϕ are bound occurrences and so are not replaced. The fourth occurrence, however, is free and so we replace it by $\mathbf{0}'$.

(iii) $\exists x\ ((x \cdot x) = y \ \&\ x = (z + y))$

Here all the occurrences of x are bound, so none get replaced and $\phi(\tau/x)$ is the same as ϕ.

(iv) $\exists y\ y = (y' + y')$

(v) $\exists z\ (x + z) = (x \cdot z)$

Here τ is the same as x, so replacing free occurrences of x by occurrences of τ leaves the formula ϕ unaltered. ■

In terms of the notation we have just introduced, the formal rule we are aiming at will tell us that from a formula $\forall x \phi$ we can derive $\phi(\tau/x)$.

Before we state this rule, we need first to check that it meets our requirement of logical validity. At first sight, everything seems in order. If $\forall x \phi$ is true under some interpretation, then everything being considered must have the property specified by the formula ϕ. In particular, the object denoted by τ has this property and so $\phi(\tau/x)$ is true. This argument is almost correct, but it overlooks a technical complication which we shall explain below. In most practical applications, this complication will not cause any problems, and you should not let the bother it causes get in the way of your understanding of the essentials of what we are doing. However, for the sake of accuracy we need to discuss this technical difficulty now.

When we were discussing the difference between free and bound variables, we saw that a formula expresses only properties of the free variables occurring in it; the bound variables are just 'dummies' whose role is, for example, to express generality. The deduction from $\forall x\phi$ to $\phi(\tau/x)$ will be valid only if $\phi(\tau/x)$ says the same thing about τ as ϕ says about x, and things will go wrong if there are variables in τ which become bound when we replace free occurrences of x by τ in ϕ.

Look back at what happened in Problem 1(iv) above. In that example, $\forall x\phi$ was the formula

$$\forall x\ \exists y\ y = (x + x), \tag{1}$$

and $\phi(\tau/x)$ the formula

$$\exists y\ y = (y' + y'). \tag{2}$$

The term τ contains the variable y, and the occurrences of y in τ become bound when we replace x by τ in ϕ. So we claim that things have gone wrong, for the following reason. It is not difficult to see that whereas formula (1) is true when we interpret $+$ as the usual addition of natural numbers (under this interpretation it says that for each number there is another number which is twice the original number), formula (2) is false (where y' is interpreted as the result of adding 1 to y). Since there is an interpretation which makes (1) true but (2) false, we cannot validly deduce (2) from (1).

We see from this that the inference from $\forall x\phi$ to $\phi(\tau/x)$ is valid only provided no variable in τ becomes bound when we replace free occurrences of x by τ in ϕ. It is convenient to have a shorthand expression to express this condition and so we make the following definition.

DEFINITION We say that the term τ *may be freely substituted for the variable* x in the formula ϕ if no variable in τ becomes bound when we replace free occurrences of x in ϕ by occurrences of τ. ▲

EXAMPLE 2

(i) Let ϕ be $\exists y\ (y \cdot y) = (x \cdot (x + y))$ and let τ be $(z'' + \mathbf{0})$. The only variable in τ is z, and since there are no quantifiers involving z in ϕ, this variable certainly does not become bound when we replace x by τ in ϕ. So τ may be freely substituted for x in ϕ.

(ii) Let ϕ be $\exists z\ (z \cdot y) = (x \cdot (x + y))$, and let τ be $(z'' + \mathbf{0})$. τ may not be freely substituted for x in ϕ, since the variable z occurring in τ becomes bound in the formula $\phi(\tau/x)$, which is $\exists z\ (z \cdot y) = ((z'' + \mathbf{0}) \cdot ((z'' + \mathbf{0}) + y))$.

(iii) Let ϕ be $(x = \mathbf{0}' \vee \exists x\ y = x')$ and let τ be $(x + y)$. $\phi(\tau/x)$ is $((x + y) = \mathbf{0}' \vee \exists x\ y = x')$—remember we replace only free occurrences of x—and both the variables in τ, x and y, are free in $\phi(\tau/x)$ at the occurrences which arise from the replacement of x by τ. So τ may be freely substituted for x in ϕ.

However, τ may not be freely substituted for y in ϕ. $\phi(\tau/y)$ is the formula $(x = \mathbf{0}' \vee \exists x\ (x + y) = x')$ and the x in the term $(x + y)$ is bound at the place where y has been replaced by τ.

PROBLEM 2 In which of the following cases may the term τ be freely substituted for the variable x in the formula ϕ?

(i) ϕ is $(\exists z\ z = y'\ \&\ \exists t\ (t \cdot t) = x)$; τ is $(z'' \cdot y)$.

(ii) ϕ is $\exists z\ (z = y'\ \&\ \exists t\ (t \cdot t) = x)$; τ is $(z'' \cdot y)$.

(iii) ϕ is $(\forall y\ \exists z\ y = (z + z) \rightarrow (\exists t\ (t + t) = x \vee x = z))$; τ is $(y + x)$.

(iv) ϕ is $\forall y\ \exists x\ ((x + x) = y \vee (x + x') = y)$; τ is $(y + x)$.

(v) ϕ is $\forall y\ (\exists x\ (x + x) = y \rightarrow y = (x' + x))$; τ is x.

TION 2 (i) τ may be freely substituted for x in ϕ. The formula $\phi(\tau/x)$ is

$$(\exists z\ z = y' \ \&\ \exists t\ (t \cdot t) = \underbrace{(z'' \cdot y)}_{\tau})$$

and the variables y, z occurring in τ remain free when x is replaced by τ.

(ii) τ may not be freely substituted for x in ϕ. $\phi(\tau/x)$ is

$$\exists z\ (z = y' \ \&\ \exists t\ (t \cdot t) = (\underset{\underset{\text{bound}}{\uparrow}}{z''} \cdot y))$$

and the variable z in τ becomes bound when x is replaced by τ.

(iii) τ may be freely substituted for x in ϕ. $\phi(\tau/x)$ is

$$(\forall y\ \exists z\ y = (z + z) \rightarrow (\exists t\ (t + t) = \underbrace{(y + x)}_{\tau} \vee \underbrace{(y + x)}_{\tau} = z))$$

and both the variables x, y occurring in τ remain free when the two free occurrences of x in ϕ are replaced by τ.

(iv) τ may be freely substituted for x in ϕ. Notice that there are no free occurrences of x in ϕ. So $\phi(\tau/x)$ is identical with ϕ and, in a vacuous way, none of the variables in τ become bound when we replace the free occurrences of x in ϕ—there are none—by τ.

(v) τ may be freely substituted for x in ϕ. Since τ is the same as x, $\phi(\tau/x)$ is the same as ϕ, so no problem arises. ■

We can now state our formal rule for the elimination of the universal quantifier. Since this rule applies whatever the quantified variable, in stating the rule we use v to stand for any of the variables of our language.

RULE *Universal Quantifier Elimination Rule* (*UE*)
If the formula $\forall v \phi$ occurs on a certain line in a formal proof and τ is a term which may be freely substituted for the variable v in ϕ, then on any subsequent line we may derive the formula $\phi(\tau/v)$, and this formula will depend on the same assumptions as did $\forall v \phi$. ▲

AMPLE 3 We cannot give any very interesting examples of the use of the rule UE until we have some other rules for the quantifiers at our disposal. Meanwhile, here is a fairly simple example of a formal proof which uses this rule twice.

1	(1)	$\forall x\ \forall y\ (x + y) = (y + x)$	Ass
1	(2)	$\forall y\ ((z + x) + y) = (y + (z + x))$	UE, (1)
1	(3)	$((z + x) + x) = (x + (z + x))$	UE, (2)

If we call the formula on line (1) $\forall x \phi$, so that ϕ is the formula $\forall y\ (x + y) = (y + x)$, then the formula on line (2) is $\phi((z + x)/x)$, and the use of UE is legitimate since the term $(z + x)$ may be freely substituted for x in ϕ.

If we call the formula on line (2) $\forall y \psi$, so that ψ is the formula $((z + x) + y) = (y + (z + x))$, then the formula on line (3) is $\psi(x/y)$ and again the use of rule UE is legitimate since the term x may be freely substituted for y in ψ. Indeed, as there are no quantifiers in ψ, any term may be freely substituted for y in ψ.

ROBLEM 3 Show that the formula $(\mathbf{0}' + \mathbf{0}'') = (\mathbf{0}'' + \mathbf{0}')$ can be derived from the assumption $\forall x\ \forall y\ (x + y) = (y + x)$.

LUTION 3

1	(1)	$\forall x\ \forall y\ (x + y) = (y + x)$	Ass
1	(2)	$\forall y\ (\mathbf{0}' + y) = (y + \mathbf{0}')$	UE, (1)
1	(3)	$(\mathbf{0}' + \mathbf{0}'') = (\mathbf{0}'' + \mathbf{0}')$	UE, (2) ■

6.3 The Universal Quantifier Introduction Rule

To find the appropriate rule for introducing universal quantifiers into proofs, we need to consider how universal statements are proved in mathematics.

Consider, for example, the proof of the following theorem about groups. (Don't worry if you are not familiar with group theory; the mathematical details of the proof are not important for our purposes.)

Theorem Let G be a group. Then

$$\text{for all } g, h \text{ in } G, \quad (gh)^{-1} = h^{-1}g^{-1}. \qquad (*)$$

Proof Take g, h in G. $(h^{-1}g^{-1})(gh) = h^{-1}(g^{-1}g)h = h^{-1}eh = h^{-1}h = e$, where e is the identity element of G. Thus $h^{-1}g^{-1}$ is the inverse of gh, i.e.

$$(gh)^{-1} = h^{-1}g^{-1}. \qquad (\dagger)$$

This completes the proof. ■

This proof establishes the universal statement $(*)$ about all elements of the group G. We proved it by taking two elements g, h of G and showing that they satisfy the formula $(\dagger)$. Why does this enable us to claim that we have thereby proved the universal statement $(*)$ and hence that we have completed the proof of the theorem? The reason is that in deriving $(\dagger)$ we made no special assumptions about g and h. We used several properties of groups (without mentioning them explicitly) such as the associative law, but all we assumed about g and h was that they were elements of G. That is why the universal statement $(*)$ follows from $(\dagger)$.

You may have met the same idea in proofs of theorems in geometry. For example, to prove that every isosceles triangle has equal base angles, we take an isosceles triangle ABC and, making no further assumptions about it, deduce that its base angles are equal. Because we make no other assumptions about the triangle ABC, this entitles us to say that *every* isosceles triangle has equal base angles.

Thus if we want to prove a universal statement $\forall v\phi$, the standard method is to prove that ϕ is true without making special assumptions about v. It then follows that ϕ is true for every v being considered, that is, that $\forall v\phi$ is true.

We saw earlier that a formula tells us something about v if v occurs in it as a free variable. Thus in deriving ϕ, none of the formulas used as assumptions must contain v as a free variable. This leads us to the following rule.

RULE *Universal Quantifier Introduction Rule (UI)*
If the formula ϕ occurs on a line of a formal proof and the variable v does not occur free in any of the assumptions in force on this line, then on any subsequent line we may derive $\forall v\phi$, and this formula will depend on the same assumptions as did ϕ. ▲

EXAMPLE 1

1	(1)	$\forall x\,\forall y\,\forall z\,(x + (y + z)) = ((x + y) + z)$	Ass
1	(2)	$\forall y\,\forall z\,(x + (y + z)) = ((x + y) + z)$	UE, (1)
1	(3)	$\forall z\,(x + (x + z)) = ((x + x) + z)$	UE, (2)
1	(4)	$(x + (x + x)) = ((x + x) + x)$	UE, (2)
1	(5)	$\forall x\,(x + (x + x)) = ((x + x) + x)$	UI, (4)

This example is typical, in the sense that we made several applications of the UE rule before we were ready to apply the UI rule. You will see that in this example each time we used the UE rule we replaced the variable which is being eliminated by x.

To see whether the use of UI on line (5) is legitimate, we need to look at the line from which line (5) is derived. The annotation at the right of this line tells us that line (5) is derived from line (4). So we need to look at those formulas which are in force as assumptions on line (4). In the example above, the only formula used as an assumption to derive line (4) is the formula on line (1). We see that the variable x has no free occurrences in this formula. So we may legitimately use the rule UI to derive from line (4) the formula obtained by prefixing $\forall x$ to the formula on that line. The rule UI then tells us that line (5) depends on exactly the same assumptions as does line (4). Note that x has free occurrences in the formula on line (4) but, as this formula is not used as an assumption, this fact is irrelevant.

XAMPLE 2

1	(1)	$\forall x\ \forall y\ (x + y) = (y + x)$	Ass
1	(2)	$\forall y\ (x + y) = (y + x)$	UE, (1)
1	(3)	$(x + y) = (y + x)$	UE, (2)
1	(4)	$\forall x\ (x + y) = (y + x)$	UI, (3)
1	(5)	$\forall y\ \forall x\ (x + y) = (y + x)$	UI, (4)

On line (4) we use the rule UI to add the prefix $\forall x$ to the formula on line (3). So to see whether this is a legitimate use of UI we must check that the variable x does not occur free in the assumption formulas on which line (3) depends (here just the single formula on line (1)). Similarly, to see whether the use of UI on line (5) is legitimate, we must check that the variable y does not occur free in the formulas used as assumptions to derive line (4). Since, in fact, there are no free variables in the formula on line (1), it is easily seen that both uses of UI in this proof are legitimate.

ROBLEM 1 Show that the formula $\forall x\ ((x + x)\cdot x) = ((x\cdot x) + (x\cdot x))$ can be derived from the assumption $\forall x\ \forall y\ \forall z\ ((x + y)\cdot z) = ((x\cdot z) + (y\cdot z))$.

ROBLEM 2 Show that the formula $\forall x\ (x + \mathbf{0}) = (\mathbf{0} + x)$ can be derived from the assumption $\forall x\ \forall y\ (x + y) = (y + x)$.

LUTION 1

1	(1)	$\forall x\ \forall y\ \forall z\ ((x + y)\cdot z) = ((x\cdot z) + (y\cdot z))$	Ass
1	(2)	$\forall y\ \forall z\ ((x + y)\cdot z) = ((x\cdot z) + (y\cdot z))$	UE, (1)
1	(3)	$\forall z\ ((x + x)\cdot z) = ((x\cdot z) + (x\cdot z)$	UE, (2)
1	(4)	$((x + x)\cdot x) = ((x\cdot x) + (x\cdot x))$	UE, (3)
1	(5)	$\forall x\ ((x + x)\cdot x) = ((x\cdot x) + (x\cdot x))$	UI, (4) ■

LUTION 2

1	(1)	$\forall x\ \forall y\ (x + y) = (y + x)$	Ass
1	(2)	$\forall y\ (x + y) = (y + x)$	UE, (1)
1	(3)	$(x + \mathbf{0}) = (\mathbf{0} + x)$	UE, (2)
1	(4)	$\forall x\ (x + \mathbf{0}) = (\mathbf{0} + x)$	UI, (3)

Note that we cannot use the rule UE to go directly from line (1) to line (4) since the rule UE enables us only to eliminate a quantifier which occurs at the very beginning of a formula. ■

OTATION It is convenient at this stage to introduce the following notation. We write

$$\phi_1, \ldots, \phi_k \vdash \psi$$

to mean that the formula ψ can be derived from the assumptions $\phi_1, \ldots, \phi_k$, that is, that there is a formal proof on whose last line the formula ψ occurs and the assumptions in force on this line are included among the formulas $\phi_1, \ldots, \phi_k$. The symbol $\vdash$ is often called the *turnstile* symbol and was introduced into logic by Gottlob Frege.

Thus, in terms of this new notation, the formal proof in Example 2 shows that $\forall x\ \forall y\ (x + y) = (y + x) \quad \vdash \quad \forall y\ \forall x\ (x + y) = (y + x)$. It is not difficult to see that this is a particular instance of the fact that

for every formula ϕ and all variables u, v,

$$\forall u\ \forall v\ \phi \ \vdash \forall v\ \forall u\ \phi.$$

In the cassette tape section which follows, we give some more examples of general results involving the derivability of formulas, and some hints on how to discover proofs.

START THE TAPE WHEN YOU ARE READY

Glossary of Pronunciation

&	$\rightarrow$	$\forall$	$\vdash$	ϕ	ψ	θ	τ
'and'	'implies'	'universal'	'derives'	'phi'	'psi'	'theta'	'tau'

Universal Quantifier Elimination Rule (*UE*)
If the formula $\forall v\phi$ occurs on a certain line in a formal proof where v is a variable and τ is a term which may be freely substituted for v in ϕ, then on any subsequent line we may derive the formula $\phi(\tau/v)$, and this formula will depend on the same assumptions as did $\forall v\phi$.

Special case of UE when τ is the same as v: from $\forall v\phi$ we may derive ϕ, and this formula will depend on the same assumptions as did $\forall v\phi$.

Universal Quantifier Introduction Rule (*UI*)
If the formula ϕ occurs on a line of a formal proof and the variable v does not occur free in any of the assumptions in force on this line, then on any subsequent line we may derive $\forall v\phi$, and this formula will depend on the same assumptions as did ϕ.

EXAMPLE 3 For all formulas ϕ, ψ, $\quad \forall x\phi, \forall x(\phi \rightarrow \psi) \ \vdash \ \forall x\psi$.

1	(1)	$\forall x\phi$	Ass
2	(2)	$\forall x(\phi \rightarrow \psi)$	Ass
1	(3)	ϕ	UE, (1)
2	(4)	$(\phi \rightarrow \psi)$	UE, (2)
1, 2	(5)	ψ	Taut, (3), (4)
1, 2	(6)	$\forall x\psi$	UI, (5)

Tautology used in line (5):

((ϕ	&	(ϕ	$\rightarrow$	ψ))	$\rightarrow$	ψ)
1	1	1	1	1	1	1
1	0	1	0	0	1	0
0	0	0	1	1	1	1
0	0	0	1	0	1	0

↑
tautology

BLEM 3 Show that for all formulas ϕ, ψ, θ,

$$\forall x(\phi \to \psi), \forall x(\psi \to \theta) \vdash \forall x(\phi \to \theta).$$

LUTION 3

1	(1)	$\forall x(\phi \to \psi)$	Ass
2	(2)	$\forall x(\psi \to \theta)$	Ass
1	(3)	$(\phi \to \psi)$	UE, (1)
2	(4)	$(\psi \to \theta)$	UE, (2)
1, 2	(5)	$(\phi \to \theta)$	Taut, (3), (4)
1, 2	(6)	$\forall x(\phi \to \theta)$	UI, (5)

Tautology used in line (5): $(((\phi \to \psi) \,\&\, (\psi \to \theta)) \to (\phi \to \theta))$ ■

MPLE 4 For all formulas ϕ, ψ, $\forall x(\phi \to \psi) \vdash (\forall x\phi \to \forall x\psi)$.

Outline Proof

1	(1)	$\forall x(\phi \to \psi)$	Ass
1	(2)	$(\phi \to \psi)$	UE, (1)
3	(3)	$\forall x\phi$	Ass
⋮	⋮	⋮	
1, 3	(k)	$\forall x\psi$	
1	(k + 1)	$(\forall x\phi \to \forall x\psi)$	CP, (k)

Strategy:
To derive $(A \to B)$:
assume A,
derive B,
use CP to get $(A \to B)$.

Full Proof

1	(1)	$\forall x(\phi \to \psi)$	Ass
1	(2)	$(\phi \to \psi)$	UE, (1)
3	(3)	$\forall x\phi$	Ass
3	(4)	ϕ	UE, (3)
1, 3	(5)	ψ	Taut, (2), (4)
1, 3	(6)	$\forall x\psi$	UI, (5)
1	(7)	$(\forall x\phi \to \forall x\psi)$	CP, (6)

Tautology used in line (5): $((\phi \,\&\, (\phi \to \psi)) \to \psi)$

BLEM 4 Show that for all formulas ϕ, ψ, $\forall x(\phi \,\&\, \psi) \vdash (\forall x\phi \,\&\, \forall x\psi)$.

BLEM 5 Show that for all formulas ϕ, ψ, θ,

$$\forall x(\phi \to (\psi \,\&\, \theta)) \vdash (\forall x\phi \to (\forall x\psi \,\&\, \forall x\theta)).$$

TION 4 *Outline Proof*

1	(1)	$\forall x(\phi \,\&\, \psi)$	Ass
1	(2)	$(\phi \,\&\, \psi)$	UE, (1)
⋮	⋮	⋮	
1	(j)	$\forall x\phi$	
⋮	⋮	⋮	
1	(k)	$\forall x\psi$	
1	(k + 1)	$(\forall x\phi \,\&\, \forall x\psi)$	Taut, (j), (k)

Tautology used in line (k + 1): $((\forall x\phi \,\&\, \forall x\psi) \to (\forall x\phi \,\&\, \forall x\psi))$

Full Proof

1	(1)	$\forall x(\phi \,\&\, \psi)$	Ass
1	(2)	$(\phi \,\&\, \psi)$	UE, (1)
1	(3)	ϕ	Taut, (2)
1	(4)	$\forall x\phi$	UI, (3)
1	(5)	ψ	Taut, (2)
1	(6)	$\forall x\psi$	UI, (5)
1	(7)	$(\forall x\phi \,\&\, \forall x\psi)$	Taut, (4), (6)

Tautologies used in line (3): $((\phi \,\&\, \psi) \rightarrow \phi)$

line (5): $((\phi \,\&\, \psi) \rightarrow \psi)$

line (7): $((\forall x\phi \,\&\, \forall x\psi) \rightarrow (\forall x\phi \,\&\, \forall x\psi))$ ■

SOLUTION 5

1	(1)	$\forall x(\phi \rightarrow (\psi \,\&\, \theta))$	Ass
1	(2)	$(\phi \rightarrow (\psi \,\&\, \theta))$	UE, (1)
3	(3)	$\forall x\phi$	Ass
3	(4)	ϕ	UE, (3)
1, 3	(5)	ψ	Taut, (2), (4)
1, 3	(6)	$\forall x\psi$	UI, (5)
1, 3	(7)	θ	Taut, (2), (4)
1, 3	(8)	$\forall x\theta$	UI, (7)
1, 3	(9)	$(\forall x\psi \,\&\, \forall x\theta)$	Taut, (6), (8)
1	(10)	$(\forall x\phi \rightarrow (\forall x\psi \,\&\, \forall x\theta))$	CP, (9)

Tautologies used in line (5): $(((\phi \rightarrow (\psi \,\&\, \theta)) \,\&\, \phi) \rightarrow \psi)$

line (7): $(((\phi \rightarrow (\psi \,\&\, \theta)) \,\&\, \phi) \rightarrow \theta)$

line (9): $((\forall x\psi \,\&\, \forall x\theta) \rightarrow (\forall x\psi \,\&\, \forall x\theta))$ ■

Some guidelines for finding formal proofs can be found in the Appendix of this Unit.

6.4 The Existential Quantifier Introduction Rule

In informal mathematics the direct method for proving that there exists an object with a certain property is to exhibit an example of an object which has that property. Here is a simple example.

Theorem There is a non-zero 2×2 matrix whose square is the zero 2×2 matrix.

Proof Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$. Then $\mathbf{A}^2 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Hence $\mathbf{A}$ is not the zero 2×2 matrix but $\mathbf{A}^2$ is the zero 2×2 matrix. This completes the proof. ■

The pattern of this sort of argument gives us our formal rule for existential quantifier introduction. If we have a formula $\phi(\tau/x)$ which expresses the fact that a term τ has a certain property, then we can derive $\exists x\phi$ which expresses the fact that there is some object which has that property.

Put the other way round, this says that to derive $\exists x\phi$, the direct method is to find a term τ such that when we replace free occurrences of x in ϕ by τ to obtain the formula $\phi(\tau/x)$ we can then derive the formula $\phi(\tau/x)$. From this we can then deduce $\exists x\phi$. As with the rule UE, we need for technical reasons to impose the restriction that τ may be freely substituted for x in ϕ.

Thus we are led to the following rule.

RULE *Existential Quantifier Introduction Rule* (EI)
If τ is a term which may be freely substituted for the variable v in the formula ϕ, and the formula $\phi(\tau/v)$ occurs on a line of a formal proof, then on any subsequent line we may derive the formula $\exists v\phi$, and this formula will depend on the same assumptions as does $\phi(\tau/v)$. ▲

XAMPLE 1 We show that $\exists y\ \mathbf{0}'' = (y + y) \ \vdash\ \exists z\ \exists y\ z = (y + y)$.

1	(1)	$\exists y\ \mathbf{0}'' = (y + y)$	Ass
1	(2)	$\exists z\ \exists y\ z = (y + y)$	EI, (1)

If we let ϕ be the formula $\exists y\ z = (y + y)$, then the formula on line (1) is $\phi(\mathbf{0}''/z)$ and the formula on line (2) is $\exists z\phi$. Since the term $\mathbf{0}''$ may be freely substituted for z in ϕ, this is a legitimate use of the rule EI.

XAMPLE 2 We show that $\forall x\ (x + \mathbf{0}) = x \ \vdash\ \forall x\ \exists y\ (x + y) = x$.

1	(1)	$\forall x\ (x + \mathbf{0}) = x$	Ass
1	(2)	$x + \mathbf{0} = x$	UE, (1)
1	(3)	$\exists y\ (x + y) = x$	EI, (2)
1	(4)	$\forall x\ \exists y\ (x + y) = x$	UI, (3)

Note that we cannot use EI to go directly from line (1) to line (4) since the rule EI allows us to introduce an existential quantifier only at the very beginning of a formula.

ROBLEM 1 Show that

(i) $\forall x\ (x + \mathbf{0}) = x \ \vdash\ \exists y\ \forall x\ (x + y) = x$;
(ii) $\forall x\ (x \cdot \mathbf{0}') = x \ \vdash\ \forall x\ \exists y\ (x \cdot y) = x$;
(iii) $\forall x\ (x + \mathbf{0}) = x \ \vdash\ \exists x\ (x + x) = x$.

LUTION 1 (i)

1	(1)	$\forall x\ (x + \mathbf{0}) = x$	Ass
1	(2)	$\exists y\ \forall x\ (x + y) = x$	EI, (1)

If we let ϕ be the formula $\forall x\ (x + y) = x$, then the formula on line (1) is $\phi(\mathbf{0}/y)$ and the formula on line (2) is $\exists y\phi$. Since the term $\mathbf{0}$ may be freely substituted for y in ϕ this is a legitimate use of the rule EI. (It should be clear that if a term τ contains no variables, then τ may be freely substituted for any variable in any formula.)

(ii)

1	(1)	$\forall x\ (x \cdot \mathbf{0}') = x$	Ass
1	(2)	$(x \cdot \mathbf{0}') = x$	UE, (1)
1	(3)	$\exists y\ (x \cdot y) = x$	EI, (2)
1	(4)	$\forall x\ \exists y\ (x \cdot y) = x$	UI, (3)

(iii)

1	(1)	$\forall x\ (x + \mathbf{0}) = x$	Ass
1	(2)	$(\mathbf{0} + \mathbf{0}) = \mathbf{0}$	UE, (1)
1	(3)	$\exists x\ (x + x) = x$	EI, (2)

Notice that in deriving line (2) we are taking ϕ to be the formula $(x + \mathbf{0}) = x$ and we are using the rule UE to go from $\forall x\phi$ to $\phi(\mathbf{0}/x)$. However, when we come to derive line (3) we are taking ϕ to be the formula $(x + x) = x$ and we are using the rule EI to go from $\phi(\mathbf{0}/x)$ to $\exists x\phi$. ■

Some further examples and problems involving the use of the rule EI will be found in the next section.

6.5 The Existential Hypothesis Rule

The fourth rule of our formal system for the quantifiers is rather more subtle than the other three. By analogy with the rules for the universal quantifier, you might expect there to be an existential quantifier elimination rule which would enable us to drop an initial $\exists x$ from the beginning of a formula. But deductions of this kind would not, in general, be valid. From the fact that there is some object which has a certain property, we cannot deduce that any particular object has that property. So any rule which enabled us to derive $\phi(\tau/x)$ from $\exists x\phi$ could not possibly be logically valid.

To discover what the second existential quantifier rule should be, we need to consider how we make deductions from existential statements in informal mathematics.

The example we have chosen to look at comes from the theory of matrices. Again, do not worry if you are not familiar with this piece of mathematics since it is the logical form of the argument which we are interested in and not the mathematical details.

We shall begin by giving a sample proof, just as you might find it in an algebra textbook. Then, in order to make the logical form clearer, we shall rewrite the proof using logical symbols. Finally, we shall extract from this the logical rule we are seeking.

Here then is our sample proof.

Theorem If the matrix **A** has an inverse, then the determinant of **A** is non-zero.

Proof Suppose **B** is the inverse of **A**.

Then $\mathbf{AB} = \mathbf{I}$ and hence $\det(\mathbf{A}) \times \det(\mathbf{B}) = \det(\mathbf{AB}) = \det(\mathbf{I}) = 1$. Hence $\det(\mathbf{A}) \neq 0$.

This completes the proof. ■

Now let us symbolize '**B** is the inverse of **A**' by **B** Inv **A**. The theorem asserts that from $\exists$**B B** Inv **A** it follows that $\det(\mathbf{A}) \neq 0$, and, leaving out the algebraic details, the proof of the theorem has the following form:

Suppose **B** Inv **A**.

⋮

algebra

⋮

Hence $\det(\mathbf{A}) \neq 0$.

This completes the proof.

Although the theorem asserts that $\det(\mathbf{A}) \neq 0$ follows from the assumption that $\exists$**B B** Inv **A**, we see that in the proof we do not start from this assumption and deduce that $\det(\mathbf{A}) \neq 0$. Instead, the assumption with which we start the proof is **B** Inv **A**. However, when we finally manage to deduce that $\det(\mathbf{A}) \neq 0$, we claim that the proof has been completed, i.e. that we have, in fact, shown that $\det(\mathbf{A}) \neq 0$ follows from the assumption $\exists$**B B** Inv **A**.

The reason we can make this claim is that in the course of the proof we made no assumptions about **B** other than that given by our initial assumption **B** Inv **A**, and the conclusion we finally reach, $\det(\mathbf{A}) \neq 0$, makes no mention of **B**. So this conclusion really depends only on the assumption that there is *some* matrix **B** which is the inverse of **A**, i.e. the conclusion really depends only on the assumption $\exists$**B B** Inv **A**, which is what the theorem asserts.

This example is typical of the way existential hypotheses are used in mathematical arguments. To deduce some conclusion from the assumption that there is some object x which satisfies the property expressed by ϕ, i.e. from the assumption $\exists x\phi$, we assume that x is an object which has the property expressed by ϕ and deduce the desired conclusion without making any further assumptions about x.

So the formal rule we are looking for needs to express the fact that, if we can derive a formula ψ from the assumption ϕ and possibly other assumptions which do not contain x as a free variable, then we can also derive ψ from $\exists x\phi$ and the other assumptions. For this rule to be logically valid we must also impose the condition that ψ 'makes no mention' of x, i.e. that x does not occur as a free variable in ψ. Thus we are led to the following rule.

RULE *The Existential Hypothesis Rule* (*EH*)
Let ψ be a formula which contains no free occurrences of the variable v. If on a certain line of a formal proof we have derived the formula ψ from the assumption ϕ and possibly other assumptions in which v does not occur free (though v may occur free in ϕ), then on any subsequent line we may derive ψ from $\exists v\phi$ and these other assumptions. ▲

AMPLE 1 (*This example is also discussed in the cassette tape for this section.*)

For all formulas ϕ, θ, $\exists x\phi, \forall x(\phi \rightarrow \theta) \vdash \exists x\theta$.

1	(1)	$\exists x\phi$	Ass
2	(2)	$\forall x(\phi \rightarrow \theta)$	Ass
3	(3)	ϕ	Ass
2	(4)	$(\phi \rightarrow \theta)$	UE, (2)
2, 3	(5)	θ	Taut, (3), (4)
2, 3	(6)	$\exists x\theta$	EI, (5)
1, 2	(7)	$\exists x\theta$	EH, (6)

Tautology used in line (5): $((\phi \,\&\, (\phi \rightarrow \theta)) \rightarrow \theta)$

Since $\exists x\phi$ is one of the assumptions from which we wish to derive $\exists x\theta$, we write this formula down as an assumption on line (1) so that it gets a number. However, we make no use of line (1) until the very end of the proof. We know that, because of the rule EH, to derive $\exists x\theta$ from the assumptions $\exists x\phi$ and $\forall x(\phi \rightarrow \theta)$ it will be good enough to derive $\exists x\theta$ from the assumptions ϕ and $\forall x(\phi \rightarrow \theta)$. So we write these formulas down as assumptions on lines (2) and (3). By line (6), we have derived $\exists x\theta$ from the assumptions on lines (2) and (3). Clearly, the variable x does not occur free in $\exists x\theta$; nor does x occur free in the additional assumption, $\forall x(\phi \rightarrow \theta)$, used to derive $\exists x\theta$ on line (6). Since on this line we have derived $\exists x\theta$ from ϕ and the additional assumption $\forall x(\phi \rightarrow \theta)$, the rule EH tells us that we can derive $\exists x\theta$ from $\exists x\phi$ and the same additional assumption. This we have done on line (7).

AMPLE 2 For all formulas ϕ, ψ, $\exists x(\phi \,\&\, \psi) \vdash (\exists x\phi \,\&\, \exists x\psi)$.

1	(1)	$\exists x(\phi \,\&\, \psi)$	Ass
2	(2)	$(\phi \,\&\, \psi)$	Ass
2	(3)	ϕ	Taut, (2)
2	(4)	$\exists x\phi$	EI, (3)
2	(5)	ψ	Taut, (2)
2	(6)	$\exists x\psi$	EI, (5)
2	(7)	$(\exists x\phi \,\&\, \exists x\psi)$	Taut, (4), (6)
1	(8)	$(\exists x\phi \,\&\, \exists x\psi)$	EH, (7)

Tautologies used in line (3): $((\phi \mathbin{\&} \psi) \rightarrow \phi)$

line (5): $((\phi \mathbin{\&} \psi) \rightarrow \psi)$

line (7): $((\exists x\phi \mathbin{\&} \exists x\psi) \rightarrow (\exists x\phi \mathbin{\&} \exists x\psi))$

In this example there are no additional assumptions, so to see that the use of EH on line (8) is legitimate, we need only observe that the variable x does not occur free in the formula $(\exists x\phi \mathbin{\&} \exists x\psi)$. ☐

The following cassette-tape section looks at further examples of formal proofs using the rules EI and EH, and looks at the strategies used to produce such proofs.

START THE TAPE WHEN YOU ARE READY

Glossary of Pronunciation

&	→	–	∀	∃	⊢	ϕ	ψ	θ	τ
'and'	'implies'	'not'	'universal'	'existential'	'derives'	'phi'	'psi'	'theta'	'tau'

Existential Quantifier Introduction Rule (*EI*)
If τ is a term which may be freely substituted for the variable v in the formula ϕ, and the formula $\phi(\tau/v)$ occurs on a line of a formal proof, then on any subsequent line we may derive $\exists v\phi$, and this formula depends on the same assumptions as did $\phi(\tau/v)$.

Existential Hypothesis Rule (*EH*)
Let ψ be a formula which contains no free occurrences of the variable v. If on a certain line of a formal proof we have derived the formula ψ from the assumption ϕ and possibly other assumptions in which v does not occur free (though v may occur free in ϕ), then on any subsequent line we may derive ψ from $\exists v\phi$ and these other assumptions.

EXAMPLE 3 For all formulas ϕ, $\exists x\,\exists y\phi \vdash \exists y\,\exists x\phi$.

1	(1)	$\exists x\,\exists y\phi$	Ass
2	(2)	$\exists y\phi$	Ass
3	(3)	ϕ	Ass
3	(4)	$\exists x\phi$	EI, (3)
3	(5)	$\exists y\,\exists x\phi$	EI, (4)
2	(6)	$\exists y\,\exists x\phi$	EH, (5)
1	(7)	$\exists y\,\exists x\phi$	EH, (6)

PROBLEM 1 Show that for all formulas ϕ, ψ, where the variable x does not occur free in ψ,
$\forall x(\phi \rightarrow \psi) \vdash (\exists x\phi \rightarrow \psi)$.

SOLUTION 1

1	(1)	$\forall x(\phi \rightarrow \psi)$	Ass
1	(2)	$(\phi \rightarrow \psi)$	UE, (1)
3	(3)	$\exists x\phi$	Ass
4	(4)	ϕ	Ass
1, 4	(5)	ψ	Taut, (2), (4)
1, 3	(6)	ψ	EH, (5)
1	(7)	$(\exists x\phi \rightarrow \psi)$	CP, (6)

Tautology used in line (5): $((\phi \mathbin{\&} (\phi \rightarrow \psi)) \rightarrow \psi)$ ■

AMPLE 4 For all formulas ϕ, $\exists x -\phi \vdash -\forall x\phi$.

Outline Proof

	1	(1)	$\exists x -\phi$	Ass
	2	(2)	$-\phi$	Ass
proof by contradiction	3	(3)	$\forall x\phi$	Ass
	⋮	⋮	⋮	
	2, 3	(k)	$(\theta \,\&\, -\theta)$	
	2	$(k+1)$	$(\forall x\phi \rightarrow (\theta \,\&\, -\theta))$	CP, (k)
	2	$(k+2)$	$-\forall x\phi$	Taut, $(k+1)$
	1	$(k+3)$	$-\forall x\phi$	EH, $(k+2)$

Note: a *contradiction* is a formula of the form $(\theta \,\&\, -\theta)$.

General form of the tautology needed in proof by contradiction

$((\psi$	$\rightarrow$	$(\theta$	$\&$	$-$	$\theta))$	$\rightarrow$	$-$	$\psi)$
1	0	1	0	0	1	1	0	1
1	0	0	0	1	0	1	0	1
0	1	1	0	0	1	1	1	0
0	1	0	0	1	0	1	1	0

Full Proof

1	(1)	$\exists x -\phi$	Ass
2	(2)	$-\phi$	Ass
3	(3)	$\forall x\phi$	Ass
3	(4)	ϕ	UE, (3)
2, 3	(5)	$(\phi \,\&\, -\phi)$	Taut, (2), (4)
2	(6)	$(\forall x\phi \rightarrow (\phi \,\&\, -\phi))$	CP, (5)
2	(7)	$-\forall x\phi$	Taut, (6)
1	(8)	$-\forall x\phi$	EH, (7)

Tautologies used in line (5): $((-\phi \,\&\, \phi) \rightarrow (\phi \,\&\, -\phi))$

line (7): $((\forall x\phi \rightarrow (\phi \,\&\, -\phi)) \rightarrow -\forall x\phi)$

ROBLEM 2 Show that for all formulas ϕ, $\exists x\phi \vdash -\forall x -\phi$.

ROBLEM 3 Show that for all formulas ϕ, ψ, $\forall x(\psi \rightarrow \phi) \vdash (\exists x -\phi \rightarrow -\forall x\psi)$.

LUTION 2

1	(1)	$\exists x\phi$	Ass
2	(2)	ϕ	Ass
3	(3)	$\forall x -\phi$	Ass
3	(4)	$-\phi$	UE, (3)
2, 3	(5)	$(\phi \,\&\, -\phi)$	Taut, (2), (4)
2	(6)	$(\forall x -\phi \rightarrow (\phi \,\&\, -\phi))$	CP, (5)
2	(7)	$-\forall x -\phi$	Taut, (6)
1	(8)	$-\forall x -\phi$	EH, (7)

Tautologies used in line (5): $((\phi \,\&\, -\phi) \rightarrow (\phi \,\&\, -\phi))$

line (7): $((\forall x -\phi \rightarrow (\phi \,\&\, -\phi)) \rightarrow -\forall x -\phi)$ ■

SOLUTION 3

1	(1)	$\forall x(\psi \to \phi)$	Ass
1	(2)	$(\psi \to \phi)$	UE, (2)
3	(3)	$\exists x\ -\phi$	Ass
4	(4)	$-\phi$	Ass
5	(5)	$\forall x\psi$	Ass
5	(6)	ψ	UE, (5)
1, 5	(7)	ϕ	Taut, (2), (6)
1, 4, 5	(8)	$(\phi\ \&\ -\phi)$	Taut, (4), (7)
1, 4	(9)	$(\forall x\ \psi \to (\phi\ \&\ -\phi))$	CP, (8)
1, 4	(10)	$-\forall x\psi$	Taut, (9)
1, 3	(11)	$-\forall x\psi$	EH, (10)
1	(12)	$(\exists x\ -\phi \to -\forall x\psi)$	CP, (11)

Tautologies used in line (7): $(((\psi \to \phi)\ \&\ \psi) \to \phi)$

line (8): $((-\phi\ \&\ \phi) \to (\phi\ \&\ -\phi))$

line (10): $((\forall x\psi \to (\phi\ \&\ -\phi)) \to -\forall x\psi)$ ■

Some guidelines for finding formal proofs can be found in the Appendix of this unit.

PROBLEM 4 Show that for all formulas ϕ, ψ, $\exists x(\phi \vee \psi),\ \forall x\ -\phi\ \vdash\ \exists x\psi$.

PROBLEM 5 Show that for every formula ϕ, $\exists x\, \forall y\phi\ \vdash\ \forall y\, \exists x\phi$.

PROBLEM 6 Find the mistakes in the following 'proofs'.

(i) We show that for each formula ϕ, $\exists x\phi \vdash \forall x\phi$!!

1	(1)	$\exists x\phi$	Ass
2	(2)	ϕ	Ass
2	(3)	$\forall x\phi$	UI, (2)
1	(4)	$\forall x\phi$	EH, (3)

(ii) We show that for all formulas ϕ, ψ, $\exists x\phi,\ \exists x\psi \vdash \exists x(\phi\ \&\ \psi)$!!

1	(1)	$\exists x\phi$	Ass
2	(2)	$\exists x\psi$	Ass
3	(3)	ϕ	Ass
4	(4)	ψ	Ass
3, 4	(5)	$(\phi\ \&\ \psi)$	Taut, (3), (4)
3, 4	(6)	$\exists x(\phi\ \&\ \psi)$	EI, (5)
1, 4	(7)	$\exists x(\phi\ \&\ \psi)$	EH, (6)
1, 2	(8)	$\exists x(\phi\ \&\ \psi)$	EH, (7)

Tautology used in line (5): $((\phi\ \&\ \psi) \to (\phi\ \&\ \psi))$

We hope it is clear to you that if our formal system does meet requirement (b) of Section 6.0 that our rules are logically valid then there must be something wrong with these formal proofs, since in both cases the conclusion is not a logical consequence of the assumptions.

OLUTION 4

1	(1)	$\exists x(\phi \vee \psi)$	Ass
2	(2)	$\forall x - \phi$	Ass
3	(3)	$(\phi \vee \psi)$	Ass
2	(4)	$-\phi$	UE, (2)
2, 3	(5)	ψ	Taut, (3), (4)
2, 3	(6)	$\exists x\psi$	EI, (5)
1, 2	(7)	$\exists x\psi$	EH, (6)

Tautology used in line (5): $(((\phi \vee \psi) \,\&\, -\phi) \rightarrow \psi)$ ■

OLUTION 5

1	(1)	$\exists x\, \forall y\phi$	Ass
2	(2)	$\forall y\phi$	Ass
2	(3)	ϕ	UE, (2)
2	(4)	$\exists x\phi$	EI, (3)
2	(5)	$\forall y\, \exists x\phi$	UI, (4)
1	(6)	$\forall y\, \exists x\phi$	EH, (5)

This proof is rather pretty in that it uses each quantifier rule once. As usual when we are trying to deduce a formula from an assumption of the form $\exists x\psi$, we make the extra assumption ψ and try to derive the desired conclusion from that. The crucial steps are lines (5) and (6). The use of UI on line (5) is legitimate since the variable y does not occur free in the formula on line (2), which is the only assumption on which line (4) depends. On line (5) we have derived our desired conclusion from the assumption $\forall y\phi$. Since x does not occur free in the conclusion on line (5), we can now use EH to derive this same conclusion from the assumption $\exists x\, \forall y\phi$. ■

OLUTION 6

(i) The mistake is on line (3). In general, the variable x will have free occurrences in the formula ϕ, and line (2) has ϕ as an assumption. So the use of UI to obtain line (3) is not legitimate.

(ii) The mistake is in the use of EH on line (7). In general, the variable x will have free occurrences in the formula ψ. This formula is used as an additional assumption on line (6). So the use of EH to derive line (7) is not legitimate. Notice, on the other hand, that there is nothing wrong with the use of EH on line (8), since in this case the additional assumption on which line (7) depends, that is, the formula $\exists x\phi$ on line (1), does not contain any free occurrences of the variable x. ■

6.6 Identity Rules

There are two rules governing the identity symbol $=$: an introduction rule and a substitution rule. The introduction rule is very straightforward. It is based on the principle that everything is identical with itself. (The saying 'as sure as eggs is eggs' is thought by some to be a corruption of 'as sure as x is x', i.e. 'as sure as $x = x$'.)

RULE *Identity Introduction Rule* (*II*)
For each term τ the formula

$$\tau = \tau$$

may be written down on any line of a proof and does not depend on any assumptions. ▲

EXAMPLE 1 We show that $\vdash \forall x\ x = x$. (The absence of any formulas to the left of the turnstile, $\vdash$, indicates that there is a formal proof of which the last line is $\forall x\ x = x$, and that this line does not depend on any assumptions.)

(1)	$x = x$	II
(2)	$\forall x\ x = x$	UI, (1)

Notice that since line (1) does not depend on any assumptions, the use of UI on line (2) is legitimate since, vacuously, the variable x does not occur free in any of the assumptions—there are none—on which line (1) depends. ☐

We have to be a little more careful with the other rule for $=$. It corresponds to the principle that 'equals can be substituted for equals'. That is, if two terms τ_1, τ_2 have been shown to be equal, in the sense that we have derived the formula $\tau_1 = \tau_2$, then we should be able to substitute one for the other in any formula. There is, however, in transferring this principle into our formal system, a difficulty of the same kind as we met with the rule UE. For consider the following example.

$$\exists z\ z = (y + \mathbf{0}') \tag{1}$$

$$y = (z + \mathbf{0}') \tag{2}$$

$$\exists z\ z = ((z + \mathbf{0}') + \mathbf{0}') \tag{3}$$

If we interpret $+$ as being addition of natural numbers, $\mathbf{0}'$ as the number 1, free occurrences of y as referring to the number 2, and free occurrences of z as referring to the number 1, then formulas (1) and (2) are true, but (3) is false. (Notice that z has no free occurrences in (3).) We have obtained formula (3) from formula (1) by substituting the term $(z + \mathbf{0}')$ for the term y, using the fact that formula (2) says that these terms are equal. Since this substitution has led from true formulas (under our given interpretation) to a false formula, it does not correspond to a valid rule of inference. Clearly, what has gone wrong is that when we make the substitution, the occurrence of z that we introduce has become bound in the resulting formula (3).

To avoid this difficulty, we formulate our rule as given below. Although the formulation is rather long-winded in order to deal with the problem just mentioned, its use in practice is usually straightforward.

RULE *Substitution Rule (SR)*
Suppose that each of the terms τ_1 and τ_2 may be freely substituted for the variable v in the formula ϕ. Let ϕ_1 be a formula which results when some (but not necessarily all) of the free occurrences of v in ϕ are replaced by τ_1, and let ϕ_2 be the formula that results when these same occurrences are replaced by τ_2. Then if on two lines of a formal proof the formulas $\tau_1 = \tau_2$ and ϕ_1 occur, on any subsequent line we may derive ϕ_2, and ϕ_2 will then depend on all the assumptions in force on the lines on which the formulas $\tau_1 = \tau_2$ and ϕ_1 occur. ▲

Let us give some examples of formal proofs that use these rules.

EXAMPLE 2 $\vdash \forall x\ \forall y(x = y \rightarrow x' = y')$

1	(1)	$x = y$	Ass
	(2)	$x' = x'$	II
1	(3)	$x' = y'$	SR, (1), (2)
	(4)	$(x = y \rightarrow x' = y')$	CP, (3)
	(5)	$\forall y(x = y \rightarrow x' = y')$	UI, (4)
	(6)	$\forall x\ \forall y(x = y \rightarrow x' = y')$	UI, (5)

In terms of the notation used in stating the rule SR, ϕ is the formula $x' = v'$, τ_1 is x and τ_2 is y. Thus the formula on line (1) is $\tau_1 = \tau_2$, the formula on line (2) is ϕ_1 which is $\phi(x/v)$, and the formula on line (3) is ϕ_2 which is $\phi(y/v)$. However, it is probably easier to think of line (3) as arising from the replacement of the second occurrence of x in the formula in line (2) by y. This substitution is justified by the rule SR since we have the formula $x = y$ on line (1).

EXAMPLE 3 $\vdash \forall x\, \forall y(x = y \rightarrow y = x)$

1	(1)	$x = y$	Ass
	(2)	$x = x$	II
1	(3)	$y = x$	SR, (1), (2)
	(4)	$(x = y \rightarrow y = x)$	CP, (3)
	(5)	$\forall y(x = y \rightarrow y = x)$	UI, (4)
	(6)	$\forall x\, \forall y(x = y \rightarrow y = x)$	UI, (5)

Line (3) here is justified by the rule SR since the formula on line (3) has been obtained by substituting y for the first occurrence of x in the formula on line (2), and $x = y$ is the formula on line (1).

PROBLEM 1 Establish the following.

(i) $\vdash \forall x\, \forall y\, \forall z(x = y \rightarrow (z + x) = (z + y))$

(ii) $\vdash \forall x\, \forall y\, \forall z(x = y \rightarrow (y = z \rightarrow x = z))$

OLUTION 1 (i)

1	(1)	$x = y$	Ass
	(2)	$(z + x) = (z + x)$	II
1	(3)	$(z + x) = (z + y)$	SR, (1), (2)
	(4)	$(x = y \rightarrow (z + x) = (z + y))$	CP, (3)
	(5)	$\forall z(x = y \rightarrow (z + x) = (z + y))$	UI, (4)
	(6)	$\forall y\, \forall z(x = y \rightarrow (z + x) = (z + y))$	UI, (5)
	(7)	$\forall x\, \forall y\, \forall z(x = y \rightarrow (z + x) = (z + y))$	UI, (6)

Here line (3) has been obtained by substituting y for the second occurrence of x in the formula on line (2).

(ii)

1	(1)	$x = y$	Ass
2	(2)	$y = z$	Ass
1, 2	(3)	$x = z$	SR, (2), (1)
1	(4)	$(y = z \rightarrow x = z)$	CP, (3)
	(5)	$(x = y \rightarrow (y = z \rightarrow x = z))$	CP, (4)
	(6)	$\forall z(x = y \rightarrow (y = z \rightarrow x = z))$	UI, (5)
	(7)	$\forall y\, \forall z(x = y \rightarrow (y = z \rightarrow x = z))$	UI, (6)
	(8)	$\forall x\, \forall y\, \forall z(x = y \rightarrow (y = z \rightarrow x = z))$	UI, (7)

Notice that here it is the formula on line (1) that plays the role of ϕ_1 and the formula on line (2) that plays the role of $\tau_1 = \tau_2$. Thus the formula on line (3) has been obtained by substituting z for y in the formula on line (1). ■

6.7 The Correctness and Adequacy of Our System

We do not have the space to justify fully that our formal system satisfies the two conditions given in Section 6.0. Nor have we developed all the technical machinery that would be needed to do so. Nevertheless, we hope that our informal discussions make it fairly clear that they are satisfied.

Consider first the requirement that there should be an algorithm to check whether a given sequence of formulas is a formal proof. To do this we have to be able to check whether the uses of the formal rules are legitimate. This involves answering such questions as 'does the variable x occur free in any of the assumptions?', and it should be fairly obvious that there is an algorithm which answers such questions.

The second condition, that the rules are logically valid, is more of a problem since we have not given any precise definition of an 'interpretation'. We hope, however, that our informal discussions have convinced you of the plausibility of the following theorem, which we merely state but which will prove to be very useful in the next unit.

The Correctness Theorem
If $\phi_1, \ldots, \phi_k \vdash \psi$, then ψ is a logical consequence of the formulas $\phi_1, \ldots, \phi_k$, i.e. any interpretation of the non-logical symbols which makes each of $\phi_1, \ldots, \phi_k$ true also makes ψ true.

We can also ask the converse question. If ψ is a logical consequence of $\phi_1, \ldots, \phi_k$, is there a formal proof of our system which demonstrates this, i.e. is it true that $\phi_1, \ldots, \phi_k \vdash \psi$? This is a deep question, which amounts to asking: is our formal system adequate for deriving all logical consequences which can be expressed in the system?

The answer to this question is 'yes', but the proof is not easy and so we omit it entirely. Indeed, although the system of formal logic described in *Units* 5 and 6 was first published (in a superficially different but essentially the same form) by Gottlob Frege in 1879, it was some 50 years before Kurt Gödel, in 1930, proved that the answer to our question is 'yes'. So we state, without any attempt at a proof, the following theorem.

The Adequacy Theorem
If ψ is a logical consequence of the formulas $\phi_1, \ldots, \phi_k$, then $\phi_1, \ldots, \phi_k \vdash \psi$.

Despite this theorem, there are many formulas that you might expect to be derivable in our system which, in fact, have no formal proofs. One such is $\forall x\ (x \cdot \mathbf{0}') = x$. The reason why this 'true' formula is not derivable, is that its truth is not a matter of pure logic alone; it depends on our interpretation of the arithmetical symbols $\cdot$ and $\mathbf{0}'$. So to be able to derive this formula we need to add some arithmetical axioms which express the basic properties of the arithmetical symbols. We turn to this task in the next unit.

6.8 Objectives of Unit 6

We list those things on which we may set assessment questions to test your understanding of the unit. After working through the unit, we feel you should be able to:

(i) determine whether occurrences of variables in formulas are free or bound;

(ii) determine whether or not a given term may be freely substituted for a given variable in a given formula;

(iii) understand the meaning and use of each of the rules UE, UI, EI, EH, II and SR;

(iv) check a purported formal proof to see whether the uses of these rules are legitimate or not;

(v) construct simple formal proofs using these rules together with the rules introduced in *Unit 5*;

(vi) explain the Correctness and Adequacy Theorems.

6.9 Appendix: Guidelines for finding formal proofs

The following tips will usually be helpful in finding formal proofs.

(1) Begin by writing down the assumption formulas.

(2) Use UE to eliminate quantifiers from formulas which begin with universal quantifiers.

(3) When assumptions begin with existential quantifiers, prepare for a use of EH by introducing a new assumption.

(4) Look at the conclusion you are trying to derive.

(5) If the conclusion has the form $\forall v\phi$, try to derive ϕ and then use UI to get $\forall v\phi$.

(6) If the conclusion has the form $(\phi \rightarrow \psi)$, introduce ϕ as a new assumption and try to derive ψ. Then use CP.

(7) If the conclusion has the form $(\phi \;\&\; \psi)$, try to derive ϕ, ψ separately and then use the Tautology Rule to get $(\phi \;\&\; \psi)$.

(8) If the conclusion has the form $-\phi$, introduce ϕ as a new assumption and try to derive a contradiction, i.e. a formula of the form $(\theta \;\&\; -\theta)$; this is often done by separately deriving θ and $-\theta$, and then deriving $(\theta \;\&\; -\theta)$ by the Tautology Rule.

(9) Look for possible uses of the Tautology Rule.

UNIT 7
FORMAL NUMBER THEORY

7.0 Introduction

In *Units 5* and *6* we described a set of rules for carrying out logical deductions. We claimed (although we omitted the proof) that these rules are adequate for deriving all logical consequences expressible in our language. However, we also remarked that there are 'true' sentences which cannot be derived.

This is not surprising since the logical rules provide only a framework for making deductions. To do mathematics within this framework we need also some *axioms* which express the basic properties of the mathematical concepts we are working with.

Our aim is to carry out deductions in number theory. That is why we included in our language the arithmetical symbols, $\mathbf{0}$, $'$, $+$, $\cdot$.

We therefore begin this unit by describing what it means for a sentence of our formal language to 'express' a mathematical property and, in particular, a property of numbers. (Recall from *Unit 6* that a sentence is defined as a formula with no free variables.) We can then explain what we mean by an axiom and then formulate the particular system for number theory which we shall be using in the rest of this course.

7.1 An Axiom System for Number Theory

We begin this section by discussing the interpretation of the arithmetical symbols.

It should be clear how we intend the arithmetical symbols to be interpreted. $\mathbf{0}$ is interpreted as standing for the number 0. $'$ stands for the successor operation of adding one, and $+$, $\cdot$ stand for addition and multiplication respectively. There is one other thing we need to specify, namely, which collection of numbers we are talking about. For consider the sentence

$$\forall x\, \exists y\, x = y'$$

which is interpreted as:

> For every number x, there is a number y such that x is obtained from y by adding one.

We cannot determine whether or not this sentence is true until we are told which set of numbers is under consideration. It is true for the set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ of all integers (since every integer has an immediate predecessor), but not true for the set $N = \{0, 1, 2, \ldots\}$ of natural numbers, since 0 is not the successor of any element in N. We see from this example that for each interpretation we have to specify a *domain* of discourse, i.e. a set of objects to which our formulas refer, or, to be precise, to which the variables in our formulas refer. If our domain is a set U then $\forall x$ is interpreted as 'for all x in U' and $\exists x$ as 'there is at least one x in U'.

The tradition in formal number theory is to regard the set N of natural numbers as the domain we are primarily interested in. We therefore call the interpretation we get by taking this set as the domain the *standard interpretation* (the arithmetical symbols being interpreted as described above). Any other interpretation of the arithmetical symbols with a specified domain is called a *non-standard interpretation*. Thus the set U of all integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ with, for instance, the same interpretation of $\mathbf{0}$, $'$, $+$, $\cdot$ as above is a non-standard interpretation.

Notice that the sentence $\forall x\ \exists y\ x = y'$ is false in the standard interpretation but true in the interpretation U of all integers just described, so it is perfectly possible for the same sentence to be true in one interpretation but false in another. Such sentences of our formal language are called *contingent*. Another example of a contingent sentence is

$$\exists y\ (y \cdot y) = \mathbf{0}''. \qquad (*)$$

In the standard interpretation this is interpreted as

> There is a natural number y, which, when multiplied by itself, yields the successor of the successor of 0,

or in other words,

> There is a natural number which is a square root of the number 2,

and this is clearly false.

Thus the sentence $(*)$ is interpreted as being false in the standard interpretation. We claim we can now find an interpretation in which $(*)$ is true. Unfortunately, the interpretation U with domain the set of all integers described above will not do, since $(*)$ is also false in this interpretation (because 2 has no square root in the set of integers). However, suppose we take as our domain the set of real numbers and interpret $\mathbf{0}$ as the real number 0, $'$ as the function 'add one', and $+, \cdot$ as addition and multiplication of real numbers, respectively. Then the sentence $(*)$ is interpreted as

> There is a real number y, which, when multiplied by itself, yields 0 plus one plus one,

or in other words

> There is a real number which is a square root of 2,

and this is true. Thus the sentence$(*)$ comes out true in this interpretation.

Thus, as the sentence $\exists y\,(y \cdot y) = \mathbf{0}''$ is true in one interpretation but false in another (namely the standard interpretation), it is contingent.

We shall continue our discussion of non-standard interpretations in Section 7.3. For the moment, our concern is the standard interpretation and the set of all sentences of our formal language which are true in the standard interpretation. We have already remarked that there are such true sentences which cannot be derived in our formal system, and this is not surprising since none of our rules make any reference to this particular interpretation of the arithmetical symbols—indeed, we made sure that our rules were logically valid, that is, valid in all possible interpretations. We must now pay more attention to the intended (i.e. standard) interpretation of the arithmetical symbols and incorporate some features of it into our formal system. We shall look for properties of addition and multiplication of natural numbers which can be expressed in our formal language by some (preferably simple) set, S say, of sentences, the idea being that as many true (in the standard interpretation) sentences as possible should be derivable from S. Then the sentences in the set S will be regarded as permanent assumptions about the standard interpretation that we may introduce into a formal proof at any stage. Sentences used as assumptions in this way are normally referred to as *axioms*.

Let us introduce some further terminology. An *axiom system* T consists of a set of axioms and all the sentences derivable from these axioms using the rules of proof in *Units* 5 and 6. The *theory* of an axiom system T is the set of all sentences derivable from the axioms in T: these sentences are described as *theorems* of T.

The formal number theory we are going to look at will be the theory (in the technical sense above) of some axiom system, and so we need to choose some axioms. How should we choose these? We suggest the following criteria.

Requirements for an axiom system

(1) The axioms must be evidently true under the standard interpretation.

(2) There must be an algorithm for determining whether a formula is an axiom.

(3) We should be able to derive as many theorems as possible from the axioms.

(4) The axioms should be as simple as possible.

(1) and (2) are precise requirements which every system of axioms must satisfy. (Of course, (2) is a precise requirement since we have a precise notion of algorithm.) Requirement (2) is a mathematical way of expressing the fact that we want to be able to 'write down explicitly' the set of axioms. Requirements (3) and (4) are not so precise and to some extent conflict with one another—the simpler the axioms are, the fewer sentences one would expect to be able to derive from them. Following **BJ**, we give greater weight to (4). We shall choose to work with a very simple set of axioms, which is, however, ideally suited to our purpose of investigating general properties of systems of formal number theory. We shall see that the system we choose to work with fails badly in some ways (but not others) on requirement (3), but, subsequently, we shall discuss how the system might be strengthened to meet this requirement. The system we shall be investigating is denoted by Q. It has seven axioms.

The Axioms of the System Q

Q1 $\forall x\, \forall y\, (x' = y' \rightarrow x = y)$

Q2 $\forall x - \mathbf{0} = x'$

Q3 $\forall x\, (-x = \mathbf{0} \rightarrow \exists y\, (x = y'))$

Q4 $\forall x\, (x + \mathbf{0}) = x$

Q5 $\forall x\, \forall y\, (x + y') = (x + y)'$

Q6 $\forall x\, (x \cdot \mathbf{0}) = \mathbf{0}$

Q7 $\forall x\, \forall y\, (x \cdot y') = ((x \cdot y) + x)$

Let us check that our first requirement for an axiom system holds for Q, that is, that each of the above axioms is true in the standard interpretation.

Firstly, *Q1* is interpreted as:

> For every natural number x and for every natural number y, *if* the successor of x is the same number as the successor of y, *then* x equals y.

As the successor function $s: N \longrightarrow N$ is a one-one function, axiom *Q1* is indeed true in the standard interpretation.

Q2 is interpreted in the standard interpretation as:

> For every natural number x, it is not the case that the successor of x is 0.

Again this is evidently true.

We shall now check *Q5* and leave you to check the rest of the axioms as an exercise, in the hope that it is now clear that the axioms of Q are merely the formal expressions of familiar properties of the natural numbers.

Q5 is interpreted as:

> For every natural number x and for every natural number y, the sum of x and the successor of y is equal to the successor of the sum of x and y.

This is again true, being, you may recall from *Unit 3*, part of the recursive definition of the sum function.

Let us now consider our second requirement for an axiom system. As Q has only finitely many axioms, there is a very simple algorithm for determining whether a formula ϕ is an axiom. Just compare ϕ with each of the seven axioms of Q in turn. If ϕ is the same as one of the sentences *Q1*–*Q7*, then it is an axiom; otherwise, it is not.

We hope you agree that the set of axioms of Q is very simple, so that our fourth requirement is met, but before we demonstrate (in Section 7.3) that our third requirement is not met, we shall, in the next section, give some examples of sentences that are theorems of Q.

PROBLEM 1 Write down the properties of the standard interpretation expressed by each of the axioms *Q3*, *Q4*, *Q6* and *Q7*, and hence satisfy yourself that each of these axioms is interpreted as being true in this interpretation.

OLUTION 1 *Q3* is interpreted as:

> Every non-zero natural number x has an immediate predecessor; that is, for every non-zero natural number x, there is a natural number whose successor is x.

Q4 is interpreted as:

> For every natural number x, the sum of x and zero is x.

Q6 is interpreted as:

> For every natural number x, the product of x with zero equals zero.

Q7 is interpreted as:

> For every natural number x and for every natural number y, the product of x with the successor of y equals the sum of the product of x with y, and x.

(Again note that *Q6* and *Q7* give the recursive definition of the product function in *Unit 3*.) ■

7.2 Proofs in the System Q

In this section we shall look at some of the theorems of the system Q.

NOTATION

(i) If T is an axiom system and ϕ is a formula, then we write $\vdash_T \phi$ to mean that there are axioms $\phi_1, \ldots, \phi_n$ in the system T such that $\phi_1, \ldots, \phi_n \vdash \phi$. Note that any formula provable using no assumptions is automatically provable in any axiom system T, i.e. if $\vdash \phi$ then $\vdash_T \phi$.

(ii) We shall use $\mathbf{n}$ to abbreviate $\mathbf{0}\underbrace{''\cdots'}_{n\ \text{times}}$, so that e.g. $\mathbf{2}$ stands for $\mathbf{0}''$. Since in the standard interpretation the symbol $\mathbf{0}$ is interpreted as 0 and the symbol $'$ is interpreted as 'add one', clearly the term $\mathbf{n}$ will be interpreted as the number n.

(iii) We shall sometimes use $\tau_1 \neq \tau_2$ as an abbreviation for the formula $-\tau_1 = \tau_2$, where τ_1 and τ_2 are terms. It is, however, generally advisable not to use this abbreviation in a formal proof.

(iv) From now on we shall not write down any tautologies used in a proof underneath that proof, as these tautologies can be easily reconstructed using the information on the lines where the Tautology Rule has been used.

EXAMPLE 1 $\vdash_Q \mathbf{1} \neq \mathbf{2}$, i.e. $\vdash_Q -\mathbf{0}' = \mathbf{0}''$.

We shall, in fact, show that $\forall x\, \forall y\, (x' = y' \rightarrow x = y), \forall x - \mathbf{0} = x' \;\vdash\; -\mathbf{0}' = \mathbf{0}''$. As the assumption formulas in this proof are in Q (they are *Q1* and *Q2* respectively), this suffices, by definition of $\vdash_Q \phi$, to show that $\vdash_Q \mathbf{1} \neq \mathbf{2}$. (In Section 7.3 we shall show that any proof of $\mathbf{1} \neq \mathbf{2}$ requires some arithmetical axioms like those of Q used above. One cannot prove $\mathbf{1} \neq \mathbf{2}$ using purely the rules of *Units* 5 and 6.)

Our proof is as follows.

1	(1)	$\forall x\, \forall y\, (x' = y' \rightarrow x = y)$	Ass
2	(2)	$\forall x\, -\mathbf{0} = x'$	Ass
3	(3)	$\mathbf{0}' = \mathbf{0}''$	Ass
1	(4)	$\forall y\, (\mathbf{0}' = y' \rightarrow \mathbf{0} = y)$	UE, (1)
1	(5)	$(\mathbf{0}' = \mathbf{0}'' \rightarrow \mathbf{0} = \mathbf{0}')$	UE, (4)
1, 3	(6)	$\mathbf{0} = \mathbf{0}'$	Taut, (3), (5)
2	(7)	$-\mathbf{0} = \mathbf{0}'$	UE, (2)
1, 2, 3	(8)	$(\mathbf{0} = \mathbf{0}' \,\&\, -\mathbf{0} = \mathbf{0}')$	Taut, (6), (7)
1,2	(9)	$(\mathbf{0}' = \mathbf{0}'' \rightarrow (\mathbf{0} = \mathbf{0}' \,\&\, -\mathbf{0} = \mathbf{0}'))$	CP, (8)
1, 2	(10)	$-\mathbf{0}' = \mathbf{0}''$	Taut, (9)

The first three lines of this proof follow the guidelines described in the cassette tapes that accompanied *Unit 6*. Thus we first write down our assumptions and then observe that the formula we are trying to prove is of the form $-\phi$, so we assume ϕ on line (3) and try to deduce a contradiction. The way we have deduced the contradiction is by formalizing the following informal argument. Suppose $1 = 2$. Now $1 = 2$ implies $0 = 1$ (by reducing both sides by one), hence $0 = 1$. However, $0 \neq 1$ so we have a contradiction.

Now it is axiom *Q1* that allows us to 'reduce both sides by one', and this explains why we have applied the rule UE on lines (4) and (5) in the way shown. We are aiming at the formula $(\mathbf{0}' = \mathbf{0}'' \rightarrow \mathbf{0} = \mathbf{0}')$ (i.e. '$1 = 2$ implies $0 = 1$'), and the easiest way to obtain it is by removing the symbols $\forall x$ from *Q1* and replacing x by $\mathbf{0}$, which we have done on line (4), and then removing $\forall y$ and replacing y by $\mathbf{0}'$, which we have done on line (5). (The uses of the rule UE are valid since the terms we are substituting here, namely $\mathbf{0}$ and $\mathbf{0}'$, contain no variables so can be freely substituted in any formula.)

We now use the axiom *Q2* to obtain, on line (7), the negation of the formula on line (6), and hence reach the required contradiction by a simple use of the rule UE, namely by removing the symbols $\forall x$ from *Q2* (on line (2)) and substituting the term $\mathbf{0}$ for x.

Lines (8), (9) and (10) of the proof are just the routine completion of a proof by contradiction as described on the second cassette tape of *Unit 6*.

PROBLEM 1 Show that $\vdash_Q \mathbf{2} \neq \mathbf{3}$, i.e. $\vdash_Q -\mathbf{0}'' = \mathbf{0}'''$. (Hint: As above, only axioms *Q1* and *Q2* need to be used as assumptions, but you will need to use *Q1* twice, once to 'reduce' $\mathbf{2} = \mathbf{3}$ to $\mathbf{1} = \mathbf{2}$, and again to 'reduce' $\mathbf{1} = \mathbf{2}$ to $\mathbf{0} = \mathbf{1}$.)

PROBLEM 2 Show that $\vdash_Q \mathbf{0} \neq \mathbf{3}$, i.e. $\vdash_Q -\mathbf{0} = \mathbf{0}'''$. (You will need only axiom *Q2* here.)

PROBLEM 3 Show that $\vdash_Q \mathbf{1} \neq \mathbf{4}$, i.e. $\vdash_Q -\mathbf{0}' = \mathbf{0}''''$. (As above, only axioms *Q1* and *Q2* will be needed here.)

SOLUTION 1

1	(1)	$\forall x\, \forall y\, (x' = y' \rightarrow x = y)$	Ass
2	(2)	$\forall x\, -\mathbf{0} = x'$	Ass
3	(3)	$\mathbf{0}'' = \mathbf{0}'''$	Ass
1	(4)	$\forall y\, (\mathbf{0}'' = y' \rightarrow \mathbf{0}' = y)$	UE, (1)
1	(5)	$(\mathbf{0}'' = \mathbf{0}''' \rightarrow \mathbf{0}' = \mathbf{0}'')$	UE, (4)
1, 3	(6)	$\mathbf{0}' = \mathbf{0}''$	Taut, (3), (5)
1	(7)	$\forall y\, (\mathbf{0}' = y' \rightarrow \mathbf{0} = y)$	UE, (1)
1	(8)	$(\mathbf{0}' = \mathbf{0}'' \rightarrow \mathbf{0} = \mathbf{0}')$	UE, (7)
1, 3	(9)	$\mathbf{0} = \mathbf{0}'$	Taut, (6), (8)
2	(10)	$-\mathbf{0} = \mathbf{0}'$	UE, (2)
1, 2, 3	(11)	$(\mathbf{0} = \mathbf{0}' \,\&\, -\mathbf{0} = \mathbf{0}')$	Taut, (9), (10)
1, 2	(12)	$(\mathbf{0}'' = \mathbf{0}''' \rightarrow (\mathbf{0} = \mathbf{0}' \,\&\, -\mathbf{0} = \mathbf{0}'))$	CP, (11)
1, 2	(13)	$-\mathbf{0}'' = \mathbf{0}'''$	Taut, (12) ■

LUTION 2

1	(1)	$\forall x - \mathbf{0} = x'$	Ass
1	(2)	$-\mathbf{0} = \mathbf{0}'''$	UE, (1)

(Here we have substituted the term $\mathbf{0}''$ for x in line (1) to get line (2).) ■

LUTION 3

1	(1)	$\forall x\, \forall y\, (x' = y' \rightarrow x = y)$	Ass
2	(2)	$\forall x - \mathbf{0} = x'$	Ass
3	(3)	$\mathbf{0}' = \mathbf{0}''''$	Ass
1	(4)	$\forall y\, (\mathbf{0}' = y' \rightarrow \mathbf{0} = y)$	UE, (1)
1	(5)	$(\mathbf{0}' = \mathbf{0}'''' \rightarrow \mathbf{0} = \mathbf{0}''')$	UE, (4)
1,3	(6)	$\mathbf{0} = \mathbf{0}'''$	Taut, (3), (5)
2	(7)	$-\mathbf{0} = \mathbf{0}'''$	UE, (2)
1,2,3	(8)	$(\mathbf{0} = \mathbf{0}''' \;\&\; -\mathbf{0} = \mathbf{0}''')$	Taut, (6), (7)
1,2	(9)	$(\mathbf{0}' = \mathbf{0}'''' \rightarrow (\mathbf{0} = \mathbf{0}''' \;\&\; -\mathbf{0} = \mathbf{0}'''))$	CP, (8)
1,2	(10)	$-\mathbf{0}' = \mathbf{0}''''$	Taut, (9) ■

The examples and problems above show that Q has enough axioms for us to derive simple inequalities between numerical expressions (i.e. terms). We shall next derive some equalities between numerical expressions, the idea being to use the recursive nature of axioms *Q4*, *Q5* and *Q6*, *Q7*. (As we have mentioned, these axioms express just the recursive definitions of addition and multiplication respectively.) Thus our proofs will be formal counterparts of the calculations we did at the beginning of *Unit 3* where we used, for example, the recursive definition of addition to show $2 + 4 = 6$.

MPLE 2 $\vdash_Q (\mathbf{2} + \mathbf{1}) = \mathbf{3}$, i.e. $\vdash_Q (\mathbf{0}'' + \mathbf{0}') = \mathbf{0}'''$.

We shall, in fact show *Q4*, *Q5* $\vdash (\mathbf{0}'' + \mathbf{0}') = \mathbf{0}'''$. Here is our proof.

1	(1)	$\forall x\, (x + \mathbf{0}) = x$	Ass
2	(2)	$\forall x\, \forall y\, (x + y') = (x + y)'$	Ass
2	(3)	$\forall y\;\; (\mathbf{0}'' + y') = (\mathbf{0}'' + y)'$	UE, (2)
2	(4)	$(\mathbf{0}'' + \mathbf{0}') = (\mathbf{0}'' + \mathbf{0})'$	UE, (3)
1	(5)	$(\mathbf{0}'' + \mathbf{0}) = \mathbf{0}''$	UE, (1)
1,2	(6)	$(\mathbf{0}'' + \mathbf{0}') = \mathbf{0}'''$	SR, (4), (5)

After writing down the assumptions on lines (1) and (2) we immediately look at the formula we are trying to derive for a clue as to which terms we should substitute in these assumptions when using the rule UE. We have chosen to substitute $\mathbf{0}''$ for x and $\mathbf{0}$ for y in axiom *Q5* (i.e. assumption 2) because the resulting formula (on line (4)) 'reduces' the term $(\mathbf{0}'' + \mathbf{0}')$ (which occurs on the left-hand side of the formula we are trying to derive) to $(\mathbf{0}'' + \mathbf{0})'$. Now axiom *Q4* (i.e. assumption 1) provides us with a more useful expression for the term $(\mathbf{0}'' + \mathbf{0})$ (see line (5)), namely $\mathbf{0}''$, and line (6) is the result of replacing $(\mathbf{0}'' + \mathbf{0})$ by $\mathbf{0}''$ in line (4). (This is a correct use of the rule SR, for let ϕ be the formula $(\mathbf{0}'' + \mathbf{0}') = x'$, τ_1 the term $(\mathbf{0}'' + \mathbf{0})$ and τ_2 the term $\mathbf{0}''$. Then the formulas on lines (4) and (5) are $\phi(\tau_1/x)$ and $\tau_1 = \tau_2$ respectively, so the rule SR allows us to derive $\phi(\tau_2/x)$, which is the formula $(\mathbf{0}'' + \mathbf{0}') = \mathbf{0}'''$.) □

For formulas involving larger numbers, more reductions (i.e. uses of *Q5* or *Q7*) will be needed before we can use the first recursive equation (i.e. *Q4* or *Q6*), as the following example shows.

MPLE 3 $\vdash_Q (\mathbf{2} + \mathbf{2}) = \mathbf{4}$, i.e. $\vdash_Q (\mathbf{0}'' + \mathbf{0}'') = \mathbf{0}''''$.

Proof

1	(1)	$\forall x\, (x + \mathbf{0}) = x$	Ass
2	(2)	$\forall x\, \forall y\, (x + y') = (x + y)'$	Ass
2	(3)	$\forall y\, (\mathbf{0}'' + y') = (\mathbf{0}'' + y)'$	UE, (2)
2	(4)	$(\mathbf{0}'' + \mathbf{0}'') = (\mathbf{0}'' + \mathbf{0}')'$	UE, (3)
2	(5)	$(\mathbf{0}'' + \mathbf{0}') = (\mathbf{0}'' + \mathbf{0})'$	UE, (3)
2	(6)	$(\mathbf{0}'' + \mathbf{0}'') = (\mathbf{0}'' + \mathbf{0})''$	SR, (4), (5)
1	(7)	$(\mathbf{0}'' + \mathbf{0}) = \mathbf{0}''$	UE, (1)
1,2	(8)	$(\mathbf{0}'' + \mathbf{0}'') = \mathbf{0}''''$	SR, (6), (7)

Notice the reductions here. Line (4) expresses $(\mathbf{2}+\mathbf{2})$ in terms of $(\mathbf{2}+\mathbf{1})$ and line (5) expresses $(\mathbf{2}+\mathbf{1})$ in terms of $(\mathbf{2}+\mathbf{0})$, both using the property of axiom *Q5*. Then the property of axiom *Q4* is used in line (7) to express $(\mathbf{2}+\mathbf{0})$ in terms of $\mathbf{0}$ and $'$. Lines (6) and (8) make the necessary 'substitutions of equals for equals' using the rule SR. ■

Proofs of formulas similar to the above involving the multiplication symbol are necessarily more complicated. This is because in reducing a term using axiom *Q7* an addition symbol is introduced, and this must be eliminated using axioms *Q4* and *Q5* as above before we can make further reductions using *Q7*. Here is an example of such a proof.

EXAMPLE 4 $\vdash_Q (\mathbf{1}\cdot\mathbf{2})=\mathbf{2}$, i.e. $\vdash_Q (\mathbf{0}'\cdot\mathbf{0}'')=\mathbf{0}''$.

Proof

1	(1)	$\forall x\,(x+\mathbf{0})=x$	Ass
2	(2)	$\forall x\,\forall y\,(x+y')=(x+y)'$	Ass
3	(3)	$\forall x\,(x\cdot\mathbf{0})=\mathbf{0}$	Ass
4	(4)	$\forall x\,\forall y\,(x\cdot y')=((x\cdot y)+x)$	Ass
4	(5)	$\forall y\,(\mathbf{0}'\cdot y')=((\mathbf{0}'\cdot y)+\mathbf{0}')$	UE, (4)
4	(6)	$(\mathbf{0}'\cdot\mathbf{0}'')=((\mathbf{0}'\cdot\mathbf{0}')+\mathbf{0}')$	UE, (5)
2	(7)	$\forall y\,((\mathbf{0}'\cdot\mathbf{0}')+y')=((\mathbf{0}'\cdot\mathbf{0}')+y)'$	UE, (2)
2	(8)	$((\mathbf{0}'\cdot\mathbf{0}')+\mathbf{0}')=((\mathbf{0}'\cdot\mathbf{0}')+\mathbf{0})'$	UE, (7)
1	(9)	$((\mathbf{0}'\cdot\mathbf{0}')+\mathbf{0})=(\mathbf{0}'\cdot\mathbf{0}')$	UE, (1)
1, 2	(10)	$((\mathbf{0}'\cdot\mathbf{0}')+\mathbf{0}')=(\mathbf{0}'\cdot\mathbf{0}')'$	SR, (8), (9)
1, 2, 4	(11)	$(\mathbf{0}'\cdot\mathbf{0}'')=(\mathbf{0}'\cdot\mathbf{0}')'$	SR, (6), (10)
4	(12)	$(\mathbf{0}'\cdot\mathbf{0}')=((\mathbf{0}'\cdot\mathbf{0})+\mathbf{0}')$	UE, (5)
2	(13)	$\forall y\,((\mathbf{0}'\cdot\mathbf{0})+y')=((\mathbf{0}'\cdot\mathbf{0})+y)'$	UE, (2)
2	(14)	$((\mathbf{0}'\cdot\mathbf{0})+\mathbf{0}')=((\mathbf{0}'\cdot\mathbf{0})+\mathbf{0})'$	UE, (13)
1	(15)	$((\mathbf{0}'\cdot\mathbf{0})+\mathbf{0})=(\mathbf{0}'\cdot\mathbf{0})$	UE, (1)
1, 2	(16)	$((\mathbf{0}'\cdot\mathbf{0})+\mathbf{0}')=(\mathbf{0}'\cdot\mathbf{0})'$	SR, (14), (15)
1, 2, 4	(17)	$(\mathbf{0}'\cdot\mathbf{0}')=(\mathbf{0}'\cdot\mathbf{0})'$	SR, (12), (16)
1, 2, 4	(18)	$(\mathbf{0}'\cdot\mathbf{0}'')=(\mathbf{0}'\cdot\mathbf{0})''$	SR, (11), (17)
3	(19)	$(\mathbf{0}'\cdot\mathbf{0})=\mathbf{0}$	UE, (3)
1, 2, 3, 4	(20)	$(\mathbf{0}'\cdot\mathbf{0}'')=\mathbf{0}''$	SR, (18), (19)

Since the assumptions on lines (1), (2), (3) and (4) are all axioms of *Q* this proof does indeed show $\vdash_Q (\mathbf{1}\cdot\mathbf{2})=\mathbf{2}$.

The proof looks rather complicated, but as we stated above it can be broken down into a few basic steps. Thus, on line (6) we make the obvious reduction of the term $(\mathbf{0}'\cdot\mathbf{0}'')$ using axiom *Q7*, but this introduces an addition symbol. We eliminate this symbol in lines (7)–(11), so that on line (11) we have $\mathbf{1}\cdot\mathbf{2}$ expressed in terms of $\mathbf{1}\cdot\mathbf{1}$. We now repeat this procedure with the term $\mathbf{1}\cdot\mathbf{1}$ and express it in terms of $\mathbf{1}\cdot\mathbf{0}$ (line (17)). We now make the necessary substitution to obtain $\mathbf{1}\cdot\mathbf{2}$ expressed in terms of $\mathbf{1}\cdot\mathbf{0}$ on line (18), and complete the proof using *Q6* (the first recursive equation for multiplication) which tells us that $\mathbf{1}\cdot\mathbf{0}=\mathbf{0}$.

PROBLEM 4 Establish the following:

(i) $\vdash_Q (\mathbf{0}+\mathbf{1})=\mathbf{1}$, i.e. $\vdash_Q (\mathbf{0}+\mathbf{0}')=\mathbf{0}'$.

(ii) $\vdash_Q (\mathbf{1}+\mathbf{1})\neq\mathbf{1}$, i.e. $\vdash_Q -(\mathbf{0}'+\mathbf{0}')=\mathbf{0}'$. (Hint: You will need axioms *Q1* and *Q2* as well as *Q4* and *Q5*.)

(iii) $\vdash_Q (\mathbf{1}\cdot\mathbf{1})=\mathbf{1}$, i.e. $\vdash_Q (\mathbf{0}'\cdot\mathbf{0}')=\mathbf{0}'$.

SOLUTION 4

(i)	1	(1)	$\forall x\,(x+\mathbf{0})=x$	Ass
	2	(2)	$\forall x\,\forall y\,(x+y')=(x+y)'$	Ass
	2	(3)	$\forall y\,(\mathbf{0}+y')=(\mathbf{0}+y)'$	UE, (2)
	2	(4)	$(\mathbf{0}+\mathbf{0}')=(\mathbf{0}+\mathbf{0})'$	UE, (3)
	1	(5)	$(\mathbf{0}+\mathbf{0})=\mathbf{0}$	UE, (1)
	1, 2	(6)	$(\mathbf{0}+\mathbf{0}')=\mathbf{0}'$	SR, (4), (5)

(ii)	1	(1) $\forall x\,(x + \mathbf{0}) = x$	Ass
	2	(2) $\forall x\,\forall y\,(x + y') = (x + y)'$	Ass
	3	(3) $(\mathbf{0}' + \mathbf{0}') = \mathbf{0}'$	Ass
	2	(4) $\forall y\,(\mathbf{0}' + y') = (\mathbf{0}' + y)'$	UE, (2)
	2	(5) $(\mathbf{0}' + \mathbf{0}') = (\mathbf{0}' + \mathbf{0})'$	UE, (4)
	2, 3	(6) $(\mathbf{0}' + \mathbf{0})' = \mathbf{0}'$	SR, (3), (5)
	7	(7) $\forall x\,\forall y\,(x' = y' \rightarrow x = y)$	Ass
	7	(8) $\forall y\,((\mathbf{0}' + \mathbf{0})' = y' \rightarrow (\mathbf{0}' + \mathbf{0}) = y)$	UE, (7)
	7	(9) $((\mathbf{0}' + \mathbf{0})' = \mathbf{0}' \rightarrow (\mathbf{0}' + \mathbf{0}) = \mathbf{0})$	UE, (8)
	2, 3, 7	(10) $(\mathbf{0}' + \mathbf{0}) = \mathbf{0}$	Taut. (6). (9)
	1	(11) $(\mathbf{0}' + \mathbf{0}) = \mathbf{0}'$	UE, (1)
	1, 2, 3, 7	(12) $\mathbf{0}' = \mathbf{0}$	SR, (10), (11)
	13	(13) $\forall x - x' = \mathbf{0}$	Ass
	13	(14) $-\mathbf{0}' = \mathbf{0}$	UE, (13)
	1,2,3,7,13	(15) $(\mathbf{0}' = \mathbf{0}\ \&\ -\mathbf{0}' = \mathbf{0})$	Taut, (12), (14)
	1, 2, 7, 13	(16) $((\mathbf{0}' + \mathbf{0}') = \mathbf{0}' \rightarrow (\mathbf{0}' = \mathbf{0}\ \&\ -\mathbf{0}' = \mathbf{0}))$	CP, (15)
	1, 2, 7, 13	(17) $-(\mathbf{0}' + \mathbf{0}') = \mathbf{0}'$	Taut, (16)
(iii)	1	(1) $\forall x\,(x + \mathbf{0}) = \mathbf{0}$	Ass
	2	(2) $\forall x\,\forall y\,(x + y') = (x + y)'$	Ass
	3	(3) $\forall x\,(x \cdot \mathbf{0}) = \mathbf{0}$	Ass
	4	(4) $\forall x\,\forall y\,(x \cdot y') = ((x \cdot y) + x)$	Ass
	4	(5) $\forall y\,(\mathbf{0}' \cdot y') = ((\mathbf{0}' \cdot y) + \mathbf{0}')$	UE, (4)
	4	(6) $(\mathbf{0}' \cdot \mathbf{0}') = ((\mathbf{0}' \cdot \mathbf{0}) + \mathbf{0}')$	UE, (5)
	2	(7) $\forall y\,((\mathbf{0}' \cdot \mathbf{0}) + y') = ((\mathbf{0}' \cdot \mathbf{0}) + y)'$	UE, (2)
	2	(8) $((\mathbf{0}' \cdot \mathbf{0}) + \mathbf{0}') = ((\mathbf{0}' \cdot \mathbf{0}) + \mathbf{0})'$	UE, (7)
	1	(9) $((\mathbf{0}' \cdot \mathbf{0}) + \mathbf{0}) = (\mathbf{0}' \cdot \mathbf{0})$	UE, (1)
	1, 2	(10) $((\mathbf{0}' \cdot \mathbf{0}) + \mathbf{0}') = (\mathbf{0}' \cdot \mathbf{0})'$	SR, (8), (9)
	1, 2, 4	(11) $(\mathbf{0}' \cdot \mathbf{0}') = (\mathbf{0}' \cdot \mathbf{0})'$	SR, (6), (10)
	3	(12) $(\mathbf{0}' \cdot \mathbf{0}) = \mathbf{0}$	UE, (3)
	1, 2, 3, 4	(13) $(\mathbf{0}' \cdot \mathbf{0}') = \mathbf{0}'$	SR, (11), (12) ■

In our last examples in this section we present proofs in the system Q of formulas that involve quantifiers. Our strategy is essentially to follow the guidelines for finding proofs set out in *Unit 6*, introducing an axiom of Q as an assumption when required to make progress. Exactly which combination of axioms is needed will usually be suggested by the form of the formula we are trying to derive.

(Since we rarely have to use all seven axioms of Q in any particular proof, we do not write all of them down as assumptions at the beginning of a proof but rather introduce them as they are needed.)

XAMPLE 5 $\vdash_Q \forall x\,\exists y\,(x + y) = x$

Proof

1	(1)	$\forall x\,(x + \mathbf{0}) = x$	Ass
1	(2)	$(x + \mathbf{0}) = x$	UE, (1)
1	(3)	$\exists y\,(x + y) = x$	EI, (2)
1	(4)	$\forall x\,\exists y\,(x + y) = x$	UI, (3)

Recall that our strategy for deriving a formula of the form $\forall x\phi$ is to try first to derive the formula ϕ. So in this case we try to derive the formula $\exists y\,(x + y) = x$. The form of this formula then suggests we try to derive the formula $(x + \tau) = x$ for some term τ. Our arithmetical intuition suggests to us that the sensible choice of τ is simply $\mathbf{0}$, and indeed $(x + \mathbf{0}) = x$ is easily derivable from axiom *Q4*.

There is a useful point to notice here. Although there is no guarantee that $(x + \mathbf{0}) = x$ can be proved in Q—indeed you will soon see that there are some very simple-looking sentences true in the standard interpretation that are not theorems of Q—there would be no point in trying to derive the formula $(x + \mathbf{n}) = x$ for any non-zero value of n. This is because for such an n the formula $(x + \mathbf{n}) = x$ is false in the standard interpretation however x is interpreted, whereas the Correctness Theorem of *Unit* 6 would lead us to expect that any formula derivable from the axioms of Q is true in the standard interpretation (as the axioms of Q are themselves true in this interpretation).

PROBLEM 5 Show that $\vdash_Q \forall x \exists y\, (x \cdot y) = y$.

PROBLEM 6 Show that $\vdash_Q \forall x\, -\mathbf{0} = (x + x')$.

SOLUTION 5

1	(1)	$\forall x\, (x \cdot \mathbf{0}) = \mathbf{0}$	Ass
1	(2)	$(x \cdot \mathbf{0}) = \mathbf{0}$	UE, (1)
1	(3)	$\exists y\, (x \cdot y) = y$	EI, (2)
1	(4)	$\forall x \exists y\, (x \cdot y) = y$	UI, (3)

As in Example 5, we should like to derive the formula $(x \cdot \tau) = \tau$ for some term τ. Intuition suggests $\mathbf{0}$ as the sensible value of τ, and indeed $(x \cdot \mathbf{0}) = \mathbf{0}$ is an easy consequence of axiom *Q6*. ■

SOLUTION 6

1	(1)	$\mathbf{0} = (x + x')$	Ass
2	(2)	$\forall x\, \forall y\, (x + y') = (x + y)'$	Ass
2	(3)	$\forall y\, (x + y') = (x + y)'$	UE, (2)
2	(4)	$(x + x') = (x + x)'$	UE, (3)
1, 2	(5)	$\mathbf{0} = (x + x)'$	SR, (1), (4)
6	(6)	$\forall x\, -\mathbf{0} = x'$	Ass
6	(7)	$-\mathbf{0} = (x + x)'$	UE, (6)
1, 2, 6	(8)	$(\mathbf{0} = (x + x)' \mathbin{\&} -\mathbf{0} = (x + x)')$	Taut, (5), (7)
2, 6	(9)	$(\mathbf{0} = (x + x') \rightarrow (\mathbf{0} = (x + x)' \mathbin{\&} -\mathbf{0} = (x + x)'))$	CP, (8)
2, 6	(10)	$-\mathbf{0} = (x + x')$	Taut, (9)
2, 6	(11)	$\forall x - \mathbf{0} = (x + x')$	UI, (10)

To derive $\forall x\, -\mathbf{0} = (x + x')$, we derive $-\mathbf{0} = (x + x')$ and use the rule UI. Since this latter formula begins with a negation symbol we assume the formula that results when this symbol is removed on line (1) and deduce a contradiction (on line (8)). ■

In this section we have seen some of the sentences that are theorems of Q. By the Correctness Theorem of *Unit 6*, as all the axioms of Q are true in the standard interpretation, all the theorems of Q are also true in this interpretation. But are all the sentences which are true in the standard interpretation actually theorems of Q? In the next section we show that this is not the case.

7.3 Non-standard Interpretations of Q

As promised in Section 7.1, we now produce a sentence which is true in the standard interpretation but is not a theorem of Q. One such example is the sentence $\forall x\, x' \neq x$. Clearly this sentence is true in the standard interpretation but we shall show below that it is not the case that $\vdash_Q \forall x\, x' \neq x$. The technique we use is based on the Correctness Theorem of *Unit 6*. Applied to the present situation, this theorem tells us that if it were the case that $\vdash_Q \forall x\, x' \neq x$, then in every interpretation in which the sentences *Q1*–*Q7* are true, the sentence $\forall x\, x' \neq x$ would also be true. Thus, to show that it is not the case that $\vdash_Q \forall x\, x' \neq x$, we have only to produce an interpretation in which all the sentences *Q1*–*Q7* are true but in which the sentence $\forall x\, x' \neq x$ is false, and this we now do.

Recall from Section 7.1 that an interpretation consists of a non-empty set U, the domain, together with three specified functions defined on U which are to be the interpretations of the arithmetical symbols, $'$, $+$, $\cdot$, and a specified element of the set U which is to be the interpretation of the symbol **0**. For the present case, we take U to be the set $N^* = N \cup \{\omega\}$, that is, the natural numbers together with one new element, ω. **0** is interpreted as 0, and the interpretations of $'$, $+$ and $\cdot$ are given by the following tables.

x	0	1	2	...	n	...	ω
x'	1	2	3	...	$n+1$	...	ω

$x+y$	0	1	...	n	...	ω
0	0	1	...	n	...	ω
1	1	2	...	$1+n$	...	ω
⋮	⋮	⋮		⋮		⋮
m	m	$m+1$	...	$m+n$	...	ω
⋮	⋮	⋮		⋮		⋮
ω	ω	ω	...	ω	...	ω

$x \cdot y$	0	1	2	...	n	...	ω
0	0	0	0	...	0	...	0
1	0	1	2	...	n	...	ω
2	0	2	4	...	$2n$	...	ω
⋮	⋮	⋮	⋮		⋮		⋮
m	0	m	$2m$	...	mn	...	ω
⋮	⋮	⋮	⋮		⋮		⋮
ω	0	ω	ω	...	ω	...	ω

It is not hard to show that all the axioms of Q are true when interpreted in N^* as described in these tables. For example, consider $Q5$: $\forall x \forall y (x + y') = (x + y)'$. To show this is true in N^* we must show that for every x and y in N^* the value of $x + y'$ when computed using these tables is the same as that of $(x + y)'$. Now when x and y are interpreted as natural numbers, then $(x + y')$ and $(x + y)'$ have the same interpretation as in the standard interpretation and so are equal. Thus the only difficult case is when at least one of x or y is interpreted as ω, but then the formula $(x + y') = (x + y)'$ will still be interpreted as being true since both these terms will take the value ω. For example, if we put $x = 2$ and $y = \omega$ then

$$2 + \omega' = 2 + \omega = \omega \quad \text{and} \quad (2 + \omega)' = \omega' = \omega.$$

Now notice that $\omega' = \omega$, so that the sentence $\forall x\, x' \neq x$ is false in N^*. Thus we have shown that $\forall x\, x' \neq x$ is not a theorem of Q.

OBLEM 1 Verify that the remaining axioms of Q are true when interpreted in N^*. (To verify a sentence of the form $\forall x\ \forall y \ldots$ you need only check the cases when at least one of x or y is interpreted as ω, for reasons similar to those in the $Q5$ case discussed above.)

OBLEM 2 Let N^{**} be the interpretation with domain $N^{**} = N \cup \{\alpha, \beta\}$ where $\alpha \neq \beta$, i.e. N together with two new elements α and β, such that **0** is interpreted as 0, and the interpretations of $'$, $+$ and $\cdot$ are given by the following tables.

x	0	1	2	…	n	…	α	β
x'	1	2	3	…	$n+1$	…	α	β

$x+y$	0	1	2	…	n	…	α	β
0	0	1	2	…	n	…	β	α
1	1	2	3	…	$1+n$	…	β	α
2	2	3	4	…	$2+n$	…	β	α
⋮	⋮	⋮	⋮		⋮		⋮	⋮
m	m	$m+1$	$m+2$	…	$m+n$	…	β	α
⋮	⋮	⋮	⋮		⋮		⋮	⋮
α	α	α	α	…	α	…	β	α
β	β	β	β	…	β	…	β	α

$x \cdot y$	0	1	2	…	n	…	α	β
0	0	0	0	…	0	…	α	β
1	0	1	2	…	n	…	α	β
2	0	2	4	…	$2n$	…	α	β
⋮	⋮	⋮	⋮		⋮		⋮	⋮
m	0	m	$2m$	…	mn	…	α	β
⋮	⋮	⋮	⋮		⋮		⋮	⋮
α	0	β	β	…	β	…	β	β
β	0	α	α	…	α	…	α	α

To find the interpretation of $(x + y)$ or $(x \cdot y)$ for given x and y, look in the x row and y column of the appropriate table.

(i) Show that all the axioms of *Q* are true in N^{**}.

(ii) Show further that each of the following sentences is false in N^{**} and hence deduce that none of them is a theorem of *Q* (although they are all true in the standard interpretation).

(a) $\forall x\, (\mathbf{0} + x) = x$
(b) $\forall x\, \forall y\, (x + y) = (y + x)$
(c) $\forall x\, (\mathbf{0} \cdot x) = \mathbf{0}$
(d) $\forall x\, \forall y\, (x \cdot y) = (y \cdot x)$

SOLUTION 1 *Q1* From the table giving the interpretation of $'$ it is clear that different elements of N^* do not have the same successor under this interpretation. So *Q1* is true.

Q2 Also it is clear from the table for $'$ that 0 is not the successor of any element of N^* and so *Q2* is true.

Q3 We already know that every non-zero natural number is the successor of an element of N^*, and since $\omega' = \omega$ under this interpretation every non-zero element of N^* is the successor of an element of N^*. Hence *Q3* is true.

Q4 $\omega + 0 = \omega$ (from the table for $+$) and so axiom *Q4* is true in N^*.

Q6 $\omega \cdot 0 = 0$ (from the table for $\cdot$) and so *Q6* is true in N^*

Q7 If $y = \omega$ and $x \neq 0$, then both $x \cdot y'$ and $(x \cdot y) + x$ equal ω. If $y = \omega$ and $x = 0$ then both $x \cdot y'$ and $(x \cdot y) + x$ are 0. If $x = \omega$ then both $x \cdot y'$ and $(x \cdot y) + x$ equal ω, no matter what y is. Thus in all cases $x \cdot y' = (x \cdot y) + x$ and so *Q7* is true in N^*. ■

OLUTION 2 (i) *Q1*, *Q2* and *Q3* are true in N^{**} by arguments similar to those in Solution 1 above.

Q4 and *Q6* can be verified just by looking at the first columns of the tables for addition and multiplication.

For *Q5*, suppose $x, y \in N^{**}$. If $x, y \in N$ then, as before, $x + y' = (x + y)'$ since addition of members of N is the same here as in the standard interpretation. For the other possible values of x and y, we just inspect the tables for successor and addition to see that the following hold:

If $x = \alpha$, $y \neq \alpha$, then $x + y' = \alpha = (x + y)'$
If $x = \alpha$, $y = \alpha$, then $x + y' = \beta = (x + y)'$
If $x = \beta$, $y \neq \beta$, then $x + y' = \beta = (x + y)'$
If $x = \beta$, $y = \beta$, then $x + y' = \alpha = (x + y)'$
If $x \in N$, $y = \alpha$, then $x + y' = \beta = (x + y)'$
If $x \in N$, $y = \beta$, then $x + y' = \alpha = (x + y)'$.

Thus in all cases $x + y' = (x + y)'$ as required.

The verification of *Q7* is similar.

(ii) (a) $\forall x\,(\mathbf{0} + x) = x$ is false in N^{**} since $0 + \alpha = \beta \neq \alpha$.
(b) $\forall x \forall y\,(x + y) = (y + x)$ is false in N^{**} since $\alpha + \beta = \alpha \neq \beta = \beta + \alpha$.
(c) $\forall x\,(\mathbf{0} \cdot x) = \mathbf{0}$ is false in N^{**} since $0 \cdot \alpha = \alpha \neq 0$.
(d) $\forall x \forall y\,(x \cdot y) = (y \cdot x)$ is false in N^{**} since $\alpha \cdot \beta = \beta \neq \alpha = \beta \cdot \alpha$. ■

The technique we have used to show certain sentences are not theorems of *Q* (by means of the Correctness Theorem) can be used to show other results of interest.

For example, let us show that the axiom *Q2* is not derivable from the remaining axioms of *Q*. (Results of this sort, showing one axiom is *independent* of other axioms, are of interest when one is trying to avoid redundancy in an axiom system.) All we need to do, using the Correctness Theorem, is to find an interpretation in which *Q1*, *Q3*, *Q4*, *Q5*, *Q6* and *Q7* are true and *Q2* is false.

We consider the interpretation U with domain the set of all integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ where the arithmetical symbols have their usual interpretation on this set. Now it is easy to check that *Q1* and *Q3–Q7* are all true in this interpretation as they express familiar properties of the integers. As *Q2* is $\forall x\, \mathbf{0} \neq x'$, all we need to do to show *Q2* is false in this interpretation is to find a number k in the domain of U for which $0 \neq k'$ is false in U, i.e. $0 = k'$ is true. And as the domain of U consists of all the integers and $'$ is interpreted as 'add one', we have that $0 = k'$ for $k = -1$, so that there is indeed a number k in the domain for which $0 \neq k'$ is false. Hence *Q2* is false in this interpretation. Then by the Correctness Theorem, *Q2* cannot be derivable from the other axioms of *Q*.

ROBLEM 3 Show that axiom *Q6* is not derivable from the remaining axioms of *Q* (*Q1–Q5* and *Q7*). (Hint: Make a small modification to the interpretation of multiplication on the set N^* described earlier in this section.)

ROBLEM 4 In Section 7.2 we showed that $\vdash_Q \mathbf{1} \neq \mathbf{2}$ (Example 1). Show that it is not the case that $\vdash \mathbf{1} \neq \mathbf{2}$, i.e. that $\mathbf{1} \neq \mathbf{2}$ cannot be proved using the rules of *Units 5* and *6* from *no* assumptions. (Hint: You have only to find some interpretation in which $\mathbf{1} \neq \mathbf{2}$ is false. It is actually possible to find such an interpretation in which the domain has only one element.)

SOLUTION 3 We make one change to the interpretation N^* described in the reading passage, namely, in the bottom left-hand corner of the multiplication table we change 0 to ω, i.e. we set $\omega \cdot 0 = \omega$ rather than 0. Then this clearly interprets the sentence $\forall x\, x \cdot \mathbf{0} = \mathbf{0}$ as being false (since now $\omega \cdot 0 \neq 0$), but the remaining axioms are true in this interpretation. For the interpretations of *Q1*–*Q5* have remained unchanged from N^* (where they were true) and we have only to check that $x \cdot y' = (x \cdot y) + x$ holds for $x = \omega$ and $y = 0$ (the other cases being the same as before) which it does, both sides being equal to ω.

Thus *Q6* is not derivable from the other axioms. ■

SOLUTION 4 Consider the domain U consisting of just one element, 0 i.e. $U = \{0\}$. Define the interpretation of the arithmetical symbols by $0' = 0$, $0 + 0 = 0$, $0 \cdot 0 = 0$, and interpret the symbol $\mathbf{0}$ as 0.

Then the interpretation of $\mathbf{1}$, i.e. $\mathbf{0}'$, is $0'$ which is 0, and the interpretation of $\mathbf{2}$, i.e. $\mathbf{0}''$ is $0''$ which is $0'$ which is 0. Thus in this interpretation the sentence $\mathbf{1} = \mathbf{2}$ is true, so $\mathbf{1} \neq \mathbf{2}$ is false. Hence by the Correctness Theorem it is not the case that $\vdash \mathbf{1} \neq \mathbf{2}$. (Note that in the case of a formula ψ provable using no assumptions, i.e. $\vdash \psi$, the Correctness Theorem tells us that ψ is true under any interpretation of the non-logical symbols.)

7.4 The Representability Theorem

The examples given in the last section of sentences which are true in the standard interpretation but not derivable in the system Q show how weak a system it is from the standpoint of our third requirement of an axiom system in Section 7.1. Indeed, you may well be thinking that Q is so weak as to be useless for any serious purpose. In this section we discuss an important theorem which shows why this impression is not quite correct. The fact that this key theorem works for Q but not for any significantly weaker system shows why we chose to study this particular system of axioms.

Recall that we showed in Section 7.2 that $\vdash_Q \mathbf{1} \neq \mathbf{2}$, whereas, clearly, a one line proof using the rule II shows that $\vdash_Q \mathbf{1} = \mathbf{1}$. These are particular cases of the more general theorem that follows, the truth of which you may have suspected after working through the first set of problems in Section 7.2.

THEOREM 1 For all natural numbers i, j:

(i) if $i = j$ then $\vdash_Q \mathbf{i} = \mathbf{j}$;
(ii) if $i \neq j$ then $\vdash_Q \mathbf{i} \neq \mathbf{j}$.

PROOF (i) The same sort of one line proof as that of $\mathbf{1} = \mathbf{1}$ will work here as follows:

(1) $\mathbf{i} = \mathbf{i}$ II

Note that this proof does not, in fact, need to use any of the axioms of Q.

(ii) Suppose that we are given i and j with $i \neq j$. Then $\mathbf{i}$ is the term $\mathbf{0}\underbrace{''\cdots'}_{i\text{ times}}$ and $\mathbf{j}$ is the term $\mathbf{0}\underbrace{''\cdots'}_{j\text{ times}}$. What we shall do is give a recipe for a proof of the formula $\mathbf{i} \neq \mathbf{j}$ in the system Q. Given specific values of i and j we could then use this recipe to give a full proof of $\mathbf{i} \neq \mathbf{j}$.

Let us suppose that $i < j$. Our formal proof would then begin with the lines

1 (1) $\mathbf{0}\underbrace{''\cdots'}_{i} = \mathbf{0}\underbrace{''\cdots'}_{j}$ Ass

2 (2) $\forall x \forall y\, (x' = y' \rightarrow x = y)$ Ass

As in Example 1 and Problems 1, 2 and 3 of Section 7.2, our aim is to derive a formula of the form $\mathbf{0} = \tau'$ for some term τ and thus arrive at a contradiction with the formula $\mathbf{0} \neq \tau'$ which is an immediate consequence of axiom *Q2*.

Our recipe is thus to continue the proof by using axiom *Q1* (assumption 2 above) and the rule UE repeatedly as in, for example, Problem 1 of Section 7.2 until we arrive at the formula $\mathbf{0} = \mathbf{0}\underbrace{''\cdots'}_{j-i}$. As $i < j$, the right-hand side of this formula is of the form τ' for some term τ (explicitly $\tau = \mathbf{0}\underbrace{''\cdots'}_{j-i-1}$).

Our recipe now tells us to write the axiom $\forall x\, -\mathbf{0} = x'$ as the next line of our proof, and on the following line derive the formula $-\mathbf{0} = \tau'$ (with τ as above) using the rule UE. We can now use the Tautology Rule to derive the contradiction $(\mathbf{0} = \tau' \ \&\ -\mathbf{0} = \tau')$, and the standard use of CP and the Tautology Rule (as in Solution 1 of Section 7.2) gives us the formula $\mathbf{i} \neq \mathbf{j}$ as the last line of the proof (depending on axioms *Q1* and *Q2*).

If $i > j$ the proof is almost identical. (It differs only in the need to use the result of Example 3 of Section 6.6 of *Unit 6* that $\vdash x = y \rightarrow y = x$ to derive the formula $\mathbf{j} = \mathbf{i}$ from the initial assumption $\mathbf{i} = \mathbf{j}$: thereafter the recipe is as above, with i and j swapped throughout.) ■

You may also have suspected that other results in Section 7.2, such as $\vdash_Q (\mathbf{2} + \mathbf{1}) = \mathbf{3}$, $\vdash_Q (\mathbf{2} + \mathbf{2}) = \mathbf{4}$ and $\vdash_Q (\mathbf{1} + \mathbf{1}) \neq \mathbf{1}$, are special cases of general theorems similar to Theorem 1 above. This is indeed the case.

THEOREM 2 For all natural numbers i, j, k such that $i + j = k$

(i) $\vdash_Q (\mathbf{i} + \mathbf{j}) = \mathbf{k}$,

and (ii) $\vdash_Q \forall x\, ((\mathbf{i} + \mathbf{j}) = x \rightarrow x = \mathbf{k})$.

PROOF The proof of part (i) consists of a generalization of the technique we used in Examples 2 and 3 of Section 7.2 (just as in the proof of Theorem 1), but we omit the details, as they are rather lengthy and not particularly interesting. We do, however, give a proof of part (ii) on the basis that part (i) is true.

Let us suppose, then, that $i + j = k$ so that, by part (i), we have a formal proof in which the last line, line (l), say, contains the formula $(\mathbf{i} + \mathbf{j}) = \mathbf{k}$ depending on assumptions from Q. (In fact, only axioms *Q4* and *Q5* would be needed.)

We continue the proof as follows:

Q	(l)	$(\mathbf{i} + \mathbf{j}) = \mathbf{k}$	
$l + 1$	($l + 1$)	$(\mathbf{i} + \mathbf{j}) = x$	Ass
$Q, l + 1$	($l + 2$)	$x = \mathbf{k}$	SR, (l), ($l + 1$)
Q	($l + 3$)	$((\mathbf{i} + \mathbf{j}) = x \rightarrow x = \mathbf{k})$	CP, ($l + 2$)
Q	($l + 4$)	$\forall x\, ((\mathbf{i} + \mathbf{j}) = x \rightarrow x = \mathbf{k})$	UI, ($l + 3$)

(For convenience, we have written Q instead of the line numbers on which the sentences from Q were introduced as assumptions into the proof of $(\mathbf{i} + \mathbf{j}) = \mathbf{k}$.) ■

ROLLARY 3 If i, j, n are natural numbers such that $i + j \neq n$, then $\vdash_Q (\mathbf{i} + \mathbf{j}) \neq \mathbf{n}$.

PROOF Let $i + j = k$, so that $k \neq n$. By Theorem 2 (i) we have that $\vdash_Q (\mathbf{i} + \mathbf{j}) = \mathbf{k}$, and by Theorem 1 (ii) we have that $\vdash_Q \mathbf{k} \neq \mathbf{n}$. By using the Substitution Rule we may replace the term $\mathbf{k}$ by the term $(\mathbf{i} + \mathbf{j})$ in the formula $\mathbf{k} \neq \mathbf{n}$, to obtain $\vdash_Q (\mathbf{i} + \mathbf{j}) \neq \mathbf{n}$, as required. ■

NOTATION We introduce some new notation to enable us to view Theorem 2 in a more general light. If ϕ is a formula and $p_1, \ldots, p_n$ are natural numbers, then $\phi(\mathbf{p_1}, \ldots, \mathbf{p_n})$ is the formula obtained from ϕ by replacing all free occurrences of the variable x_1 by the term $\mathbf{p_1}$, all free occurrences of the variable x_2 by the term $\mathbf{p_2}, \ldots$, and all free occurrences of the variable x_n by the term $\mathbf{p_n}$. (As, for each i, p_i is a natural number, we use the notational convention introduced at the beginning of Section 7.2 to get $\mathbf{p_i}$ to stand for the term $\mathbf{0}\underbrace{''\cdots'}_{p_i \text{ times}}$.)

EXAMPLE 1 (i) Let ϕ be the formula $(x_1 + x_2) = x_3$. Then $\phi(\mathbf{2}, \mathbf{3}, \mathbf{5})$ is the formula $(\mathbf{2} + \mathbf{3}) = \mathbf{5}$. Also, $\phi(\mathbf{2}, \mathbf{3})$ is the formula $(\mathbf{2} + \mathbf{3}) = x_3$.

(ii) Let ϕ be the formula $(\exists x_2\ x_1 = (x_2 + x_2)\ \&\ x_2 = \mathbf{0})$. Then $\phi(\mathbf{2}, \mathbf{3})$ is the formula $(\exists x_2\ \mathbf{2} = (x_2 + x_2)\ \&\ \mathbf{3} = \mathbf{0})$, since we must replace the occurrence of x_1 (which is free) by the term **2**, and all free occurrences of x_2 by **3**. Only the last occurrence of x_2 is free and thus gets replaced by **3**: the first three occurrences of x_2 are bound occurrences and so do not get replaced.

PROBLEM 1 (i) Let ϕ be the formula $((x_1 \cdot x_2) = x_3 \vee x_3 = x_2)$. Write out in full the following formulas: (a) $\phi(\mathbf{3}, \mathbf{4}, \mathbf{5})$, (b) $\phi(\mathbf{1}, \mathbf{2})$, (c) $\phi(\mathbf{0})$.

(ii) Let ϕ be the formula $x_2 = \mathbf{0}$. Write out in full the formulas (a) $\phi(\mathbf{1}, \mathbf{2})$, and (b) $\phi(\mathbf{3})$.

(iii) Let ϕ be the formula $(\exists x_4 (x_2 \cdot (x_1 \cdot x_3)) = x_4 \rightarrow x_4 = x_3)$. Write out in full the formulas (a) $\phi(\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4})$, and (b) $\phi(\mathbf{1}, \mathbf{2}, \mathbf{3})$.

SOLUTION 1 (i) (a) $((\mathbf{3} \cdot \mathbf{4}) = \mathbf{5} \vee \mathbf{5} = \mathbf{4})$
(b) $((\mathbf{1} \cdot \mathbf{2}) = x_3 \vee x_3 = \mathbf{2})$
(c) $((\mathbf{0} \cdot x_2) = x_3 \vee x_3 = x_2)$

(ii) (a) $\mathbf{2} = \mathbf{0}$, (b) $x_2 = \mathbf{0}$.

(iii) (a) $(\exists x_4 (\mathbf{2} \cdot (\mathbf{1} \cdot \mathbf{3})) = x_4 \rightarrow \mathbf{4} = \mathbf{3})$
(b) $(\exists x_4 (\mathbf{2} \cdot (\mathbf{1} \cdot \mathbf{3})) = x_4 \rightarrow x_4 = \mathbf{3})$ ■

We can now restate Theorem 2 using our new notation, as follows. Let ϕ be the formula $(x_1 + x_2) = x_3$. If $i + j = k$ then (i) $\vdash_Q \phi(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and (ii) $\vdash_Q \forall x_3\ (\phi(\mathbf{i}, \mathbf{j}, x_3) \rightarrow x_3 = \mathbf{k})$. The formal proof in (i) mirrors the fact that for each pair of natural numbers i and j there is some natural number k for which $i + j = k$, and the formal proof in (ii) mirrors the fact that this k is unique, so these two proofs together mirror the fact that addition is indeed a function (with domain $N \times N$ and codomain N) giving a unique value for each pair of values for its two arguments. A consequence of (i) and (ii) is that if i, j and k are natural numbers, we can prove $(\mathbf{i} + \mathbf{j}) = \mathbf{k}$ in the system Q exactly when $i + j = k$. We describe this by saying that *the formula* ϕ *represents the addition function in the system* Q (the latter part of this phrase coming from the fact that the formal proofs in (i) and (ii) are in the system Q).

This gives rise to the following important definition.

DEFINITION Let f be a function of n arguments, the arguments and values of f being natural numbers.

The function f is *representable* in Q if there is a formula ϕ in which no variables other than $x_1, \ldots, x_n, x_{n+1}$ occur free such that if $p_1, \ldots, p_n$ and j are natural numbers with $f(p_1, \ldots, p_n) = j$, then

(i) $\vdash_Q \phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{j})$

and (ii) $\vdash_Q \forall x_{n+1}\ (\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j})$.

In this case we say that the formula ϕ *represents* the function f in Q. As with the addition function, if the formula ϕ represents the function f in Q, then for all natural numbers $p_1, \ldots, p_n$ and j, the formula $\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{j})$ is provable in Q exactly when $f(p_1, \ldots, p_n) = j$.

EXAMPLE 2 As we have mentioned, Theorem 2 states that the formula $(x_1 + x_2) = x_3$ represents the addition function in Q. Thus the addition function is representable in Q.

EXAMPLE 3 In Section 7.2 we showed that $\vdash_Q (\mathbf{1} \cdot \mathbf{2}) = \mathbf{2}$ and $\vdash_Q (\mathbf{1} \cdot \mathbf{1}) = \mathbf{1}$. These are special cases of the fact that if i, j, k are natural numbers and $i \cdot j = k$ then $\vdash_Q (\mathbf{i} \cdot \mathbf{j}) = \mathbf{k}$. Again we omit the rather tedious proof (which is just a generalization of the proofs of the particular cases above). However, in the same way as part (ii) of Theorem 2 follows from part (i), we could now show that $\vdash_Q \forall x_3((\mathbf{i} \cdot \mathbf{j}) = x_3 \rightarrow x_3 = \mathbf{k})$ (assuming $i \cdot j = k$). These two results show that the formula $(x_1 \cdot x_2) = x_3$ represents the multiplication function in Q, so this function is representable in Q.

PROBLEM 2 Show that the formula $x_2 = \mathbf{0}$ represents in Q the zero function z given by $z(p_1) = 0$ for all $p_1 \in N$.

PROBLEM 3 Write down a formula which represents in Q the identity function id, given by $\mathrm{id}(p_1) = p_1$ for all $p_1 \in N$, and prove that it does represent id in this system.

PROBLEM 4 Show that if ϕ represents the function f in Q and $f(p_1, \ldots, p_n) = j$ and $k \neq j$ then $\vdash_Q -\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$. (Hint: First show that

$$-\mathbf{k} = \mathbf{j},\ \forall x_{n+1}(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j}) \vdash -\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k}),$$

then use Theorem 1 (ii) and the definition of representability.)

OLUTION 2 We must show that if ϕ is the formula $x_2 = \mathbf{0}$ and $z(p_1) = j$ then

(i) $\vdash_Q \phi(\mathbf{p_1}, \mathbf{j})$

and (ii) $\vdash_Q \forall x_2(\phi(\mathbf{p_1}, x_2) \rightarrow x_2 = \mathbf{j})$.

Now if $z(p_1) = j$, then $j = 0$ (regardless of which number p_1 is) so $\phi(\mathbf{p_1}, \mathbf{j})$ is the formula $\mathbf{j} = \mathbf{0}$, i.e. $\mathbf{0} = \mathbf{0}$. Also $\phi(\mathbf{p_1}, x_2)$ is the formula $x_2 = \mathbf{0}$, so we must show that (i) $\vdash_Q \mathbf{0} = \mathbf{0}$ and (ii) $\vdash_Q \forall x_2(x_2 = \mathbf{0} \rightarrow x_2 = \mathbf{0})$. The proofs are as follows:

(i)	(1)	$\mathbf{0} = \mathbf{0}$	II
(ii) 1	(1)	$x_2 = \mathbf{0}$	Ass
	(2)	$(x_2 = \mathbf{0} \rightarrow x_2 = \mathbf{0})$	CP, (1)
	(3)	$\forall x_2(x_2 = \mathbf{0} \rightarrow x_2 = \mathbf{0})$	UI, (2) ■

OLUTION 3 The function id is represented by the formula $x_1 = x_2$. We must show that if $\mathrm{id}(p_1) = j$ then

(i) $\vdash_Q \mathbf{p_1} = \mathbf{j}$,

and (ii) $\vdash_Q \forall x_2(\mathbf{p_1} = x_2 \rightarrow x_2 = \mathbf{j})$.

Now if $\mathrm{id}(p_1) = j$ then, by definition of id, $p_1 = j$, hence (i) follows from Theorem 1 part (i), and for (ii) we must show $\vdash_Q \forall x_2(\mathbf{p_1} = x_2 \rightarrow x_2 = \mathbf{p_1})$ which we do as follows:

1	(1)	$\mathbf{p_1} = x_2$	Ass
	(2)	$\mathbf{p_1} = \mathbf{p_1}$	II
1	(3)	$x_2 = \mathbf{p_1}$	SR, (1), (2)
	(4)	$(\mathbf{p_1} = x_2 \rightarrow x_2 = \mathbf{p_1})$	CP, (3)
	(5)	$\forall x_2(\mathbf{p_1} = x_2 \rightarrow x_2 = \mathbf{p_1})$	UI, (4)

Note that the function id could have been perfectly correctly represented by other formulas, e.g. $x_2 = x_1$ or $(x_1 = x_2\ \&\ x_1 = x_1)$. In general, a function may be represented by more than one formula. ■

OLUTION 4 As suggested by the hint, we consider the following proof:

1	(1)	$-\mathbf{k} = \mathbf{j}$	Ass
2	(2)	$\forall x_{n+1}(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j})$	Ass
3	(3)	$\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$	Ass
2	(4)	$(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k}) \rightarrow \mathbf{k} = \mathbf{j})$	UE, (2)
2, 3	(5)	$\mathbf{k} = \mathbf{j}$	Taut, (3), (4)
1, 2, 3	(6)	$(\mathbf{k} = \mathbf{j}\ \&\ -\mathbf{k} = \mathbf{j})$	Taut, (1), (5)
1, 2	(7)	$(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k}) \rightarrow (\mathbf{k} = \mathbf{j}\ \&\ -\mathbf{k} = \mathbf{j}))$	CP, (6)
1, 2	(8)	$-\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$	Taut, (7)

Now since $k \neq j$ it follows from Theorem 1 (ii) that $\vdash_Q -\mathbf{k} = \mathbf{j}$. Also $\vdash_Q \forall x_{n+1}(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j})$ since ϕ represents f in Q and $f(p_1, \ldots, p_n) = j$. Thus both $-\mathbf{k} = \mathbf{j}$ and $\forall x_{n+1}(\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j})$ can be derived from Q, and we have just shown that $-\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$ can be derived from these two formulas (and no other assumptions). Hence $-\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$ can be derived from Q (and no other assumptions), i.e. $\vdash_Q -\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{k})$. ■

At this point it is natural to ask which functions are representable in the system *Q*. The answer is given by the following theorem, which is rather surprising in the light of the facts given in the previous section about how weak *Q* is.

THEOREM **The Representability Theorem**
Every recursive function is representable in *Q*.

The proof of this theorem is the substance of Chapter 14 of **BJ**, but we do not include study of its proof as part of this course. It is, however, the only part of the proof of Gödel's Incompleteness Theorem (to be discussed in *Unit 8*) that we ask you to take on trust. (The reasons why we omit the details of the proof are essentially the tedious nature of formal proofs showing that various functions are representable in *Q* and the fact that **BJ**'s argument requires showing that the class of recursive functions is the same as the class of functions obtainable from the identity, sum and product functions and the characteristic function of equality, by means of composition and regular minimization.)

Since *Q* is such a weak system, it might be thought that by adding extra axioms to those of *Q* to get a system *T*, functions besides those that are recursive would become representable in *T*. We would, of course, define a function *f* to be representable in such a system *T* in the same way as we did for the system *Q*, but with the formals proofs occurring in *T* rather than in *Q*: i.e. *f* is *representable* in *T* if there is a formula ϕ such that if $f(p_1, \ldots, p_n) = j$ then
(i) $\vdash_T \phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, \mathbf{j})$
and (ii) $\vdash_T \forall x_{n+1} (\phi(\mathbf{p_1}, \ldots, \mathbf{p_n}, x_{n+1}) \rightarrow x_{n+1} = \mathbf{j})$.

As *T* includes all the axioms of *Q*, any formal proof in the system *Q* is also a formal proof in the system *T*. Thus the formal proofs in *Q* which show that all recursive functions are representable in *Q* also serve to show that all recursive functions are representable in *T*.

However, rather surprisingly, no other functions can be representable in *T*, provided *T* satisfies the first two requirements of an axiom system for number theory given in Section 7.1. In other words, we have the following theorem.

THEOREM Suppose *T* is an axiom system for formal number theory such that
(i) every axiom of *T* is true in the standard interpretation and
(ii) there is an algorithm (i.e. a Turing machine) that decides whether or not a formula is an axiom of *T*.
Then every function which is representable in *T* is recursive.

We shall not need this theorem in *Unit 8* and we do not attempt to give a proof.

Notice that conditions (i) and (ii) in this theorem hold for $T = Q$ (as we pointed out in Section 7.1). Combining the results of this theorem and the Representability Theorem we obtain:

THEOREM A function is representable in *Q* if and only if it is recursive.

This remarkable theorem provides the connection between our work on algorithms and recursive functions in *Units 1* to *4* with that on logic and formal proofs in *Units 5* to *7*. In fact we chose *Q* as the system of axioms to study because it has just enough axioms for the Representability Theorem to be true. It also has the advantage of consisting of a finite number of axioms (seven, in fact).

7.5 Objectives of Unit 7

We list those things on which we may set assessment questions to test your understanding of the unit. After working through the unit we feel you should be able to:

(i) test whether simple sentences of the formal language are true or false in a given interpretation;

(ii) construct simple formal proofs in the system Q;

(iii) use a non-standard interpretation and the Correctness Theorem to show that a sentence is not a theorem of Q;

(iv) show that a function is representable in Q (where, as a guide, the proofs required would be easier than that required for the addition function);

(v) understand both what it means for a function to be representable in Q and the Representability Theorem.

UNIT 8

GÖDEL'S INCOMPLETENESS THEOREMS

8.0 Introduction

In this final unit of the course we return to our reading of **BJ**. We have now assembled the technical machinery needed to prove some very important theorems. These theorems are among the most profound intellectual discoveries of the twentieth century. Thus you should not find it surprising if you find this unit hard going in places. We shall ask you to read the proofs of all the key theorems so that you will be convinced of their truth. However, our emphasis will be on the *content* of the theorems rather than on the details of their proofs.

So we hope that after reading this unit you will have a good understanding of what these theorems tell us about the scope and limitations of mathematical logic and algorithmic methods. This understanding is what we shall chiefly aim to assess. For more details as to the sort of questions we might set on this unit, you should read the list of objectives at the end.

There is a cassette tape which accompanies this unit, and which reviews informally the theorems in the unit.

8.1 Gödel Numbers

Remember that one of the two main questions with which we began the course was Leibniz's Question: *Is there an algorithm for deciding which statements of number theory are true?*

Initially, the notion of 'algorithm' used in this question was left informal, but then we made it precise in three different ways, as Turing computability, abacus computability, and in terms of recursive functions. By the end of *Unit 4*, we had completed the proof that these three different precise characterizations are equivalent. We regarded this coincidence as providing strong evidence for *Church's Thesis* that we had succeeded in formulating a precise definition which was equivalent to the informal notion of an algorithm.

We can, therefore, use any one of these formal characterizations in trying to give an answer to Leibniz's question. For convenience, we shall use the idea of a recursive function introduced in *Unit 3*. There is, however, one obvious difficulty: recursive functions have natural numbers as their arguments and values, whereas Leibniz's question refers to statements of number theory. How then can we use recursive functions to answer Leibniz's question?

'Statements of number theory' will be interpreted as meaning sentences of the formal language for number theory which we introduced in *Unit 5*. To be able to bring in recursive functions, we shall need to be able to represent these expressions of our formal language by natural numbers. This representation of expressions by numbers is called *gödel numbering* and is explained in the first reading passage. It turns out that this numbering has a second, more subtle, use which we discuss in the next section.

READ BJ pages 170 and 171.

NOTES (i) *Page 170, lines 17–25*
This paragraph refers to the second use of gödel numbering which we mentioned above and will be more clearly understood after you have worked through Section 8.2.

(ii) *Page 170, lines -2, -1*
BJ's system of gödel numbering allows for languages which have more non-logical symbols than are needed for number theory. So **BJ** allows for a whole stock, $f_0^0, f_1^0, \ldots$, of symbols for constants, and stipulates that the only constant symbol of our language, **0**, is identified with f_0^0. Similarly,

for each k, **BJ** allows for a whole stock, $f_0^k, f_1^k, \ldots$, of k-ary function symbols, and identifies $'$ with f_0^1, $+$ with f_0^2 and $\cdot$ with f_1^2. Also, **BJ** allows for a whole stock of symbols representing relations between terms. Our language has just one such symbol, $=$, which **BJ** identifies with A_0^2.

(iii) *Page 171, Tables 15-1 and 15-2*

To make things clearer, we give a table showing the godel numbers (gns) of just those symbols which occur in our formal language for number theory.

Symbol	(	)	&	∨	→	↔	−	∃	∀	**0**
gn	1	2	3	39	399	3999	39999	4	49	6

Symbol	′	+	·	=	x	y	z	. . .
gn	68	688	6889	788	5	59	599	

▲

EXAMPLE 1 The formula $\exists x\, y = (x + x)$ is assigned a gödel number by writing under each symbol its number and then stringing these numbers together:

∃	x	y	=	(	x	+	x	)
4	5	59	788	1	5	688	5	2

and hence this formula has gödel number 45597881568852. ☐

Conversely, given a number it is not difficult to 'decode' it to discover which string of symbols, if any, it represents. The given number can be split up easily into blocks representing individual symbols since these blocks can only begin with the digits 1, 2, 3, 4, 5, 6 and 7 and, moreover, these digits always begin new blocks. The decoding process is illustrated by the following example.

EXAMPLE 2

4 9	5	4	5 9	3 9 9 9 9	5 9	7 8 8	1	5	6 8 8 9	5	2
∀	x	∃	y	−	y	=	(	x	·	x	)

Given the number in the top line, the decoding process shows it is the gödel number of the formula $\forall x\, \exists y\, -y = (x \cdot x)$.

PROBLEM 1 Find the gödel numbers of the following formulas.

(i) $z = ((y \cdot \mathbf{0}'') + x)$

(ii) $\forall x\, (x' = \mathbf{0}' \leftrightarrow x = \mathbf{0})$

PROBLEM 2 Determine which formulas, if any, have the following as their gödel numbers.

(i) 15997886395997886682

(ii) 19597885399678862

(iii) 495578852

OLUTION 1 (i) 5997881159688966868268852

(ii) 49515687886683999578862 ■

OLUTION 2 (i) $(z = \mathbf{0} \vee z = \mathbf{0}')$

(ii) This is not the gödel number of a formula. It begins with the block 19 which is not the gödel number of any symbol of our language.

(iii) Although the string $\forall x\, x = x)$ corresponds to the gödel number, the string is not a formula. ■

From working through these problems it should be clear that the system of gödel numbering does meet the requirements specified in the second paragraph of **BJ**, page 170. It is not the system used by Gödel in his original paper, but since it meets these three requirements it does just as well.

8.2 Gödel's Diagonal Lemma

In the standard interpretation, the formulas of our language are interpreted as expressing properties of numbers. Now that we have assigned gödel numbers to the expressions of our language, formulas which express properties of numbers can be reinterpreted as expressing properties of formulas. Consider, for example, the formula A given by

$$\exists y\ x = ((\mathbf{10}\cdot y) + \mathbf{2}).$$

In the standard interpretation, A expresses the fact that x has remainder 2 when divided by 10; in other words, that the last digit of the decimal representation of x is a 2. A gödel number has this property if and only if the formula it represents ends with the symbol). Thus we can reinterpret A as expressing 'the formula with gn x ends with the symbol)'.

In this way, a formula can be interpreted as referring to other formulas. So the possibility exists of constructing a formula which refers to itself—a mathematical equivalent of paradoxical sentences in English, like 'this sentence is false'. This 'self-reference' is a key idea in the proofs of the theorems in this unit. The next reading passage deals with the first step in the construction of a self-referring formula.

READ BJ page 172, line 1 to page 172, line 12.

EXAMPLE 1 Let A be the formula $\exists y\ x = ((\mathbf{10}\cdot y) + \mathbf{2})$ given above. Remembering that **10** is an abbreviation for $\mathbf{0}''''''''''$ and **2** is an abbreviation for $\mathbf{0}''$, we see that the gödel number of A is

$$4595788116686868686868686868686889592688668682.$$

Let us call this number n_0 for short. Then $\ulcorner A \urcorner$ is the term which represents this number in our language and thus consists of the symbol **0** followed by n_0 occurrences of the symbol $'$.

The diagonalization of A, that is the formula $\exists x\ (x = \ulcorner A \urcorner\ \&\ A)$, is

$$\exists x\ (x = \mathbf{n}_0\ \&\ \exists y\ x = ((\mathbf{10}\cdot y) + \mathbf{2})).$$

In the standard interpretation, this says that the number n_0 has the property expressed by the formula A, and this property was that of being the gödel number of a formula which ends with the symbol). Since n_0 is the godel number of A, it follows that the diagonalization of A expresses 'the formula A ends with the symbol)'. □

In this particular case, the diagonalization of A happens to be true under the standard interpretation, but in general this need not be the case. If we use instead the formula $-A$ with A as above, the diagonalization of $-A$ expresses 'the formula $-A$ does not end with the symbol)', and this is false.

In some other books, the diagonalization of A is defined by simply substituting $\ulcorner A \urcorner$ for the free occurrences of x in A to get the formula $A(\ulcorner A \urcorner)$. Although this formula looks simple, its gödel number is much harder to calculate than that of the formula in **BJ**'s definition, and this would affect subsequent proofs. However, the two definitions yield equivalent formulas. This result is contained in the following lemma. If you want some more practice in proving formulas in quantifier logic as described in *Unit 6*, try to prove the lemma yourself before reading our proof.

LEMMA A For each formula A and each number n, $\vdash (\exists x\ (x = \mathbf{n}\ \&\ A) \leftrightarrow A(\mathbf{n}))$.

PROOF [Notice that the formula $(\phi \leftrightarrow \psi)$ is a tautological consequence of the formulas $(\phi \rightarrow \psi)$ and $(\psi \rightarrow \phi)$.]

1	(1)	$\exists x\ (x = \mathbf{n}\ \&\ A)$	Ass
2	(2)	$(x = \mathbf{n}\ \&\ A)$	Ass
2	(3)	A	Taut, (2)
2	(4)	$x = \mathbf{n}$	Taut, (2)
2	(5)	$A(\mathbf{n})$	SR, (3), (4)
1	(6)	$A(\mathbf{n})$	EH, (5)
	(7)	$(\exists x\ (x = \mathbf{n}\ \&\ A) \rightarrow A(\mathbf{n}))$	CP, (6)
8	(8)	$A(\mathbf{n})$	Ass
	(9)	$\mathbf{n} = \mathbf{n}$	II
8	(10)	$(\mathbf{n} = \mathbf{n}\ \&\ A(\mathbf{n}))$	Taut, (8), (9)
8	(11)	$\exists x\ (x = \mathbf{n}\ \&\ A)$	EI, (10)
	(12)	$(A(\mathbf{n}) \rightarrow \exists x\ (x = \mathbf{n}\ \&\ A))$	CP, (11)
	(13)	$(\exists x\ (x = \mathbf{n}\ \&\ A) \leftrightarrow A(\mathbf{n}))$	Taut, (7), (12) ■

ROLLARY A If G is the diagonalization of the formula F, $\vdash (G \leftrightarrow F(\ulcorner F \urcorner))$.

PROOF Use Lemma A, letting A be F and n the gödel number of F. Then $\exists x\ (x = \mathbf{n}\ \&\ A)$ becomes G and $A(\mathbf{n})$ becomes $F(\ulcorner F \urcorner)$. ■

The next Lemma deals with a similar result of a more complicated kind, which we shall need later in this section. We include its proof for the sake of completeness, but if you are short of time you should omit reading it.

LEMMA B For all formulas A, B, G,

$$(G \leftrightarrow \exists y\ (A(\mathbf{n},y)\ \&\ B(y))),\ \forall y\ (A(\mathbf{n},y) \leftrightarrow y = \mathbf{k})\ \vdash\ (G \leftrightarrow \exists y\ (y = \mathbf{k}\ \&\ B(y))).$$

PROOF

1	(1)	$(G \leftrightarrow \exists y\ (A(\mathbf{n},y)\ \&\ B(y)))$	Ass
2	(2)	$\forall y\ (A(\mathbf{n},y) \leftrightarrow y = \mathbf{k})$	Ass
2	(3)	$(A(\mathbf{n},y) \leftrightarrow y = \mathbf{k})$	UE, (2)
4	(4)	$\exists y\ (A(\mathbf{n},y)\ \&\ B(y))$	Ass
5	(5)	$(A(\mathbf{n},y)\ \&\ B(y))$	Ass
2, 5	(6)	$(y = \mathbf{k}\ \&\ B(y))$	Taut, (3), (5)
2, 5	(7)	$\exists y\ (y = \mathbf{k}\ \&\ B(y))$	EI, (6)
2, 4	(8)	$\exists y\ (y = \mathbf{k}\ \&\ B(y))$	EH, (7)
2	(9)	$(\exists y\ (A(\mathbf{n},y)\ \&\ B(y)) \rightarrow \exists y\ (y = \mathbf{k}\ \&\ B(y)))$	CP, (8)
10	(10)	$\exists y\ (y = \mathbf{k}\ \&\ B(y))$	Ass
11	(11)	$(y = \mathbf{k}\ \&\ B(y))$	Ass
11	(12)	$B(y)$	Taut, (11)
11	(13)	$y = \mathbf{k}$	Taut, (11)
11	(14)	$B(\mathbf{k})$	SR, (12), (13)
2	(15)	$(A(\mathbf{n},\mathbf{k}) \leftrightarrow \mathbf{k} = \mathbf{k})$	UE, (2)
	(16)	$\mathbf{k} = \mathbf{k}$	II
2,11	(17)	$(A(\mathbf{n},\mathbf{k})\ \&\ B(\mathbf{k}))$	Taut, (14), (15), (16)
2,11	(18)	$\exists y\ (A(\mathbf{n},y)\ \&\ B(y))$	EI, (17)
2,10	(19)	$\exists y\ (A(\mathbf{n},y)\ \&\ B(y))$	EH, (18)
2	(20)	$(\exists y\ (y = \mathbf{k}\ \&\ B(y)) \rightarrow \exists y\ (A(\mathbf{n},y)\ \&\ B(y)))$	CP, (19)
2	(21)	$(\exists y\ (A(\mathbf{n},y)\ \&\ B(y)) \leftrightarrow \exists y\ (y = \mathbf{k}\ \&\ B(y)))$	Taut, (9), (20)
1, 2	(22)	$(G \leftrightarrow \exists y\ (y = \mathbf{k}\ \&\ B(y)))$	Taut, (1), (21) ■

READ BJ page 172, line 13 to page 172, line −7.

NOTES (i) *Page 172, line 16*

μm is an abbreviation for 'the least m such that'. So the length function, lh, is defined using minimization. This minimization is not quite in the form that we used in *Unit 3*. However, by copying the method used in *Unit 3*, Section 3.3, Problem 3, we can easily rewrite the definition of lh in standard form as follows.

Let f be the function given by $f(n, m) = (1 \dot{-} m) + ((n + 1) \dot{-} 10^m)$. Clearly f is recursive. Also

$$\begin{aligned} f(n, m) = 0 \quad &\Leftrightarrow \quad 1 \dot{-} m = 0 \quad \text{and} \quad (n+1) \dot{-} 10^m = 0 \\ &\Leftrightarrow \quad 1 \leq m \quad \text{and} \quad n + 1 \leq 10^m \\ &\Leftrightarrow \quad 0 < m \quad \text{and} \quad n < 10^m. \end{aligned}$$

So $\mathrm{lh}(n) =$ the least m such that $f(n,m) = 0$. That is $\mathrm{lh} = \mathrm{Mn}[f]$. Thus lh is recursive.

(ii) *Page 172, lines 21–24*

Another example might help. Take, for example, $m = 49$ and $n = 1879$. Then $1\mathrm{h}(n) = 4$ and so

$$m * n = m \cdot 10^{\mathrm{lh}(n)} + n = 49 \cdot 10^4 + 1879 = 490000 + 1879 = 491879. \quad \blacktriangle$$

PROBLEM 1 Show that there is a recursive function, neg, such that if n is the gödel number of a formula A, then neg(n) is the gödel number of its negation, $-A$.

PROBLEM 2 Show that there is a recursive function, conj, such that if m, n are the gödel numbers of the formulas A, B, then conj(m,n) is the gödel number of their conjunction $(A \;\&\; B)$.

SOLUTION 1

$$\begin{array}{cc} - & A \\ 39999 & n \end{array}$$

If the gödel number of A is n, the gödel number of $-A$ is $39999 * n$. So define $\mathrm{neg}(n) = 39999 * n$. As $*$ is recursive, neg is also recursive. ■

SOLUTION 2

$$\begin{array}{ccccc} (& A & \& & B &) \\ 1 & m & 3 & n & 2 \end{array}$$

If the gödel numbers of A, B are m, n, the gödel number of $(A \;\&\; B)$ is $1 * (m * (3 * (n * 2)))$. So define $\mathrm{conj}(m, n) = 1 * (m * (3 * (n * 2)))$, and as $*$ is recursive, so is conj. ■

We shall need the result of the following problem for our next reading passage from **BJ**.

PROBLEM 3 Suppose the formula $A(x_1, x_2)$ represents the function f of one argument in an axiom system T.

Show that for any number p_1, if $f(p_1) = p_2$, then $\vdash_T \forall x_2\, (A(\mathbf{p_1}, x_2) \leftrightarrow x_2 = \mathbf{p_2})$.

SOLUTION 3 From the definition of 'representable in T' in Section 7.4 of *Unit* 7, we have that if $f(p_1) = p_2$ then

(i) $\vdash_T A(\mathbf{p_1}, \mathbf{p_2})$, and

(ii) $\vdash_T \forall x_2\, (A(\mathbf{p_1}, x_2) \rightarrow x_2 = \mathbf{p_2})$.

We shall show that

$$A(\mathbf{p_1}, \mathbf{p_2}),\ \forall x_2\, (A(\mathbf{p_1}, x_2) \rightarrow x_2 = \mathbf{p_2}) \quad \vdash \quad \forall x_2\, (A(\mathbf{p_1}, x_2) \leftrightarrow x_2 = \mathbf{p_2})$$

as follows:

1	(1)	$A(\mathbf{p_1}, \mathbf{p_2})$	Ass
2	(2)	$\forall x_2\, (A(\mathbf{p_1}, x_2) \rightarrow x_2 = \mathbf{p_2})$	Ass
2	(3)	$(A(\mathbf{p_1}, x_2) \rightarrow x_2 = \mathbf{p_2})$	UE, (2)
4	(4)	$x_2 = \mathbf{p_2}$	Ass
1, 4	(5)	$A(\mathbf{p_1}, x_2)$	SR, (1), (4)
1	(6)	$(x_2 = \mathbf{p} \rightarrow A(\mathbf{p_1}, x_2))$	CP, (5)
1, 2	(7)	$(A(\mathbf{p_1}, x_2) \leftrightarrow x_2 = \mathbf{p_2})$	Taut, (3), (6)
1, 2	(8)	$\forall x_2\, (A(\mathbf{p_1}, x_2) \leftrightarrow x_2 = \mathbf{p_2})$	UI, (7)

As both the assumptions in force on line (8) of the proof are theorems of T (by hypothesis in (i) and (ii) above), we have that

$$\vdash_T \forall x_2\ (A(\mathbf{p}_1, x_2) \leftrightarrow x_2 = \mathbf{p}_2). \quad \blacksquare$$

We have seen that if the formula B (we previously used A) can be interpreted as saying that the formula with gödel number x has a certain property, then the diagonalization of B can be interpreted as saying that the formula B itself has that property. We have not yet, however, achieved self-reference since the diagonalization of B is different from the formula B itself. Gödel showed that by using a clever trick it is possible to construct a formula G which is logically equivalent to the formula $B(\ulcorner G \urcorner)$ which expresses 'the formula G has the property expressed by B'. In this sense, G expresses 'this formula has the property expressed by B' and so G refers to itself.

The lemma which embodies Gödel's idea comes in the next reading passage. Although it is rather difficult to understand the trick used to achieve self-reference, we feel you should be able to follow the proof line by line. In any case, the important thing is that you remember what the lemma says. Relatively informally the lemma and its proof are as follows.

BJ Lemma 2   The Diagonal Lemma

Let T be a theory in which the function diag is representable. (Following **BJ**, in this unit we use *theory* to mean axiom system.) Then given any formula $B(y)$ with one free variable y, we can find a corresponding sentence G such that

$$\vdash_T (G \leftrightarrow B(\ulcorner G \urcorner)),$$

i.e. such that G is equivalent to the expression resulting from substituting its own gödel number into B.

Informal Proof

(i) Define $F(x)$ to be $\exists y\ (y = \text{diag}(x)\ \&\ B(y))$. (Needless to say we require a formal equivalent of $y = \text{diag}(x)$ in a proper proof.)

(ii) Define G to be the diagonalization of F, i.e. G is $\exists x\ (x = \ulcorner F \urcorner\ \&\ F)$.

(iii) By our Corollary A, $\vdash (G \leftrightarrow F(\ulcorner F \urcorner))$, so from the definition of $F(x)$ we have that

$$\vdash (G \leftrightarrow \exists y\ (y = \text{diag}(\ulcorner F \urcorner)\ \&\ B(y))).$$

But G is the diagonalization of F, so that the y above which among other things appears in the sub-formula $y = \text{diag}(\ulcorner F \urcorner)$, can be replaced by $\ulcorner G \urcorner$. Our Lemma B permits us to conclude that $\vdash_T (G \leftrightarrow B(\ulcorner G \urcorner))$.

READ BJ page 173, line 4 to page 173, line −12.

NOTES (i) *Page 173, lines 6 and 7*
In this unit, we are interested in all axiom systems for formal arithmetic, not just the system Q. The phrase 'language of T' should thus simply be interpreted as being our language for formal arithmetic, although in practice the lemma applies to other languages.

(ii) *Page 173, line 9*
Here and later **BJ** drops brackets from formulas where it is felt that no ambiguity will arise.

(iii) *Page 173, lines 10 and 11 (first two lines of the proof)*
This follows from Problem 3 above, using x, y instead of x_1, x_2, and n, k instead of p_1, p_2.

(iv) *Page 176, lines 15–18*
By Corollary A, $\vdash G \leftrightarrow F(\ulcorner F \urcorner)$. Since F is the formula $\exists y\ (A(x, y)\ \&\ B(y))$ and the gödel number of F is n, this can be written as

$\vdash (G \leftrightarrow \exists y\,(A(\mathbf{n}, y)\ \&\ B(y)))$. Since this equivalence can be derived using no assumptions at all, it is certainly derivable in the system *T*. This is the assertion on line 18.

(v) *Page 173, lines 21–23*

Line 21 uses just the fact that diag is representable in *T*. As this representability relies on formal proofs in the theory *T*, it is at this stage of the argument that the axioms of *T* are really needed.

The equivalence on line 22 follows from those on lines 18 and 21 using our Lemma B.

An application of our Lemma A shows that

$$\vdash_T (\exists y\ (y = \mathbf{k}\ \&\ B(y)) \leftrightarrow B(\mathbf{k})).$$

This, together with the equivalence on line 22, yields the result $\vdash_T (G \leftrightarrow B(\mathbf{k}))$ on line 23 using the tautology

$$(((\theta \leftrightarrow \phi)\ \&\ (\phi \leftrightarrow \psi)) \rightarrow (\theta \leftrightarrow \psi)). \quad ▲$$

We shall begin to exploit Gödel's Diagonal Lemma in the next section.

8.3 The Answer to Leibniz's Question

We are now in a position to prove a series of remarkable and important theorems, one of which gives an answer to Leibniz's question. We first require the definitions of some concepts needed in the proof of these theorems.

DEFINITION A theory (or, indeed, any set of sentences) *T* is said to be *consistent* if there is no formula *A* such that both $\vdash_T A$ and $\vdash_T -A$.

If *T* were not consistent, so that a formula *A* as above did exist, we could use the tautology rule with the tautology $((A\ \&\ -A) \rightarrow B)$, which holds for any formula *B*, to show that $\vdash_T B$: i.e. if *T* is not consistent, every formula is a theorem of *T*. Thus if there is a formula which is not derivable in *T*, then *T* must be consistent. Conversely, if *T* is consistent, then given a formula *A* at least one of the two formulas *A* and $-A$ cannot be a theorem of *T*. Thus we have the result

> *T* is consistent if and only if there is some formula which is not a theorem of *T*.

DEFINITION A set of sentences *T* is said to be *satisfiable* if there is an interpretation under which all the sentences in *T* are true.

If *T* is satisfiable then *T* is consistent, as follows. Suppose not, so that although *T* is satisfiable there is some formula *A* for which both $\vdash_T A$ and $\vdash_T -A$. Then, by the Correctness Theorem of *Unit* 6, under the interpretation in which the sentences of *T* are true we must have that both *A* and $-A$ are true. But for any formula *A* we cannot have both *A* and $-A$ true under the same interpretation. Hence the result. (It can be shown, conversely, that if *T* is consistent then *T* is satisfiable. A proof of this is given in **BJ** Chapter 12, but it is beyond the scope of this course.)

READ BJ page 173, line -7 to page 174, line 8.

NOTES (i) *Page 173, line −5*

The symbol $\notin$ means 'does not belong to'.

(ii) *Page 173, lines −4 to −1*

Examples 1 to 3 below are intended to make the idea of a definable set clearer. $k \not R n$ means that it is not the case that kRn.

(iii) *Page 174, lines 4–6*
As the axioms used in the formal proofs needed to show that a function f is representable in S are also axioms of any extension theory T, the same proof shows that f is representable in T. ▲

EXAMPLE 1 Let $B(x)$ be the formula $(x = \mathbf{0} \vee x = \mathbf{1})$. Then $B(x)$ defines the set $\{0, 1\}$ in the theory Q. Since, by II, $\vdash_Q \mathbf{0} = \mathbf{0}$, a use of the Tautology Rule gives $\vdash_Q (\mathbf{0} = \mathbf{0} \vee \mathbf{0} = \mathbf{1})$; similarly, $\vdash_Q (\mathbf{1} = \mathbf{0} \vee \mathbf{1} = \mathbf{1})$; so we have $\vdash_Q B(\mathbf{0})$ and $\vdash_Q B(\mathbf{1})$. Thus

(i) if $k \in \{0, 1\}$ then $\vdash_Q B(\mathbf{k})$.

Conversely, suppose $k \notin \{0, 1\}$. Then $k \neq 0$ and $k \neq 1$. So by Theorem 1 of *Unit* 7, Section 7.4, $\vdash_Q \mathbf{k} \neq \mathbf{0}$ and $\vdash_Q \mathbf{k} \neq \mathbf{1}$, whence, using the Tautology Rule, $\vdash_Q -(\mathbf{k} = \mathbf{0} \vee \mathbf{k} = \mathbf{1})$, i.e. $\vdash_Q -B(\mathbf{k})$. Thus

(ii) if $k \notin \{0, 1\}$ then $\vdash_Q -B(\mathbf{k})$.

(i) and (ii) are the two facts we need in order to establish our claim that the formula $B(x)$ defines the set $\{0, 1\}$ in the theory Q.

EXAMPLE 2 Let $B(x)$ be the formula $\exists y\ y + y = x$. Then $B(x)$ defines the set $\theta = \{0, 2, 4, \ldots\}$ of even natural numbers in Q. It is not difficult to show that (i) if $k \in \theta$ then $\vdash_Q B(\mathbf{k})$, but the proof that (ii) if $k \notin \theta$ then $\vdash_Q - B(\mathbf{k})$, is rather difficult and so we omit it.

EXAMPLE 3 Let $B(x,y)$ be the formula $\exists z\ z + x = y$. $B(x,y)$ defines the relation $\leq$ in Q. Again the proof of this is difficult and so omitted. ☐

The Lemma in the next reading passage gives our first application of Gödel's Diagonal Lemma (used on **BJ** page 174, line 18 with the formula $-C(y)$ in place of $B(y)$ in the statement of the lemma). This is the key use of self-reference in the course, and the argument used in the proof of the lemma is discussed informally on the cassette tape accompanying Section 8.6.

READ BJ page 174, line 9 to page 174, line −11.

NOTE **BJ**'s argument is a proof by contradiction. To show that the set θ of gödel numbers of theorems of T is not definable in T, we assume, on the contrary, that it is defined in T by the formula $C(y)$ and then deduce that both $\vdash_T G$ and $\vdash_T -G$, which contradicts the assumption that T is consistent.

The formal proofs in T of G and $-G$ require tautologies which **BJ** does not mention explicitly. From $\vdash_T (G \leftrightarrow -C(\mathbf{k}))$ and $\vdash_T -C(\mathbf{k})$, **BJ** deduces that $\vdash_T G$. This deduction uses the tautology $(((\phi \leftrightarrow -\psi) \,\&\, -\psi) \rightarrow \phi)$. From $\vdash_T (G \leftrightarrow -C(\mathbf{k}))$ and $\vdash_T C(\mathbf{k})$, **BJ** deduces that $\vdash_T -G$. This uses the tautology $(((\phi \leftrightarrow -\psi) \,\&\, \psi) \rightarrow -\phi)$. ▲

DEFINITION In Section 4.1 of *Unit 4* we defined the characteristic function of a condition. In a similar way we define the *characteristic function, f, of a set θ* of natural numbers by

$$f(k) = \begin{cases} 0, & \text{if } k \notin \theta, \\ 1, & \text{if } k \in \theta. \end{cases}$$

(Actually this is precisely the characteristic function of the condition '$k \in \theta$'.) We say that the set θ of natural numbers is *recursive* if its characteristic function is recursive.

The first major theorem of this section tells us that certain sets are not recursive. It is an important step towards the answer to Leibniz's question.

READ BJ page 174, line −10 to page 175, line −15.

NOTES (i) *Page 175, proof of Theorem 1*

We can summarize the three steps in the proof as follows:

(a) if T is a consistent extension of Q then the set of gödel numbers of theorems of T is not definable in T;

(b) if the characteristic function of a set is representable in T, then the set is definable in T.

(c) if the characteristic function of the set of gödel numbers of theorems of T were recursive, then it would be representable in T.

Step (a) is just the content of Lemma 3 of **BJ** page 174. Step (c) follows from the Representability Theorem of *Unit* 7, Section 7.4. All, then, that there is left to do is to prove (b). **BJ** does this in the remark in parentheses on lines 11–13 of page 175. We spell out this argument in more detail as follows.

Suppose that the characteristic function, f, of the set θ is represented in T by the formula $A(x,y)$. Then if $f(k) = n$ we have, by Problem 3 of Section 8.2, that $\vdash_T \forall y\ (A(\mathbf{k}, y) \leftrightarrow y = \mathbf{n})$. Now suppose $k \in \theta$. Then $f(k) = 1$ and hence $\vdash_T \forall y\ (A(\mathbf{k}, y) \leftrightarrow y = \mathbf{1})$. So by UE, $\vdash_T (A(\mathbf{k}, \mathbf{1}) \leftrightarrow \mathbf{1} = \mathbf{1})$. But, by II, $\vdash_T \mathbf{1} = \mathbf{1}$ and so, using the tautology $(((\phi \leftrightarrow \psi)\ \&\ \psi) \rightarrow \phi)$, it follows that $\vdash_T A(\mathbf{k}, \mathbf{1})$. Thus we have shown that

(i) if $k \in \theta$, then $\vdash_T A(\mathbf{k}, \mathbf{1})$.

Next suppose $k \notin \theta$. Then $f(k) = 0$ and hence $\vdash_T \forall y\ (A(\mathbf{k},y) \leftrightarrow y = \mathbf{0})$. So by UE, $\vdash_T (A(\mathbf{k}, \mathbf{1}) \leftrightarrow \mathbf{1} = \mathbf{0})$. But, by Theorem 1 of *Unit* 7, Section 7.4, $\vdash_T \mathbf{1} \neq \mathbf{0}$ and so using the tautology $(((\phi \leftrightarrow \psi)\ \&\ -\psi) \rightarrow -\phi)$ it follows that $\vdash_T -A(\mathbf{k}, \mathbf{1})$. Thus we have also shown that

(ii) if $k \notin \theta$, then $\vdash_T - A(\mathbf{k}, \mathbf{1})$.

(i) and (ii) are the facts we need to establish in order to prove that the formula $A(x, \mathbf{1})$ defines the set θ in the theory T.

(ii) *Page 175, line −15*

All the axioms of Q are true under the standard interpretation, so Q is satisfiable. Thus Q is consistent (by the remarks following the definition of satisfiable on page 92 of this section. ▲

We are now in a position to give an answer to Leibniz's question.

DEFINITION By *arithmetic* we shall mean the set of all sentences of our formal language (i.e. using the symbols $\mathbf{0}$, $'$, $+$, $\cdot$) which are true in the standard interpretation.

READ BJ page 176, line 17 to page 176, line −8.

NOTE *Page 176, line 19*

All the sentences of arithmetic are true in the standard interpretation, so that arithmetic is satisfiable. Thus arithmetic is consistent, by the remarks following the definition of satisfiable. As all the axioms of Q are true in the standard interpretation, arithmetic is an extension of Q. ▲

BJ Theorem 3 tells us that the answer to Leibniz's question is 'no'. There is no algorithm for determining whether an arbitrary sentence of our formal language for number theory is true under the standard interpretation. Of course, in many cases we can decide that a particular sentence is true, but **BJ** Theorem 3 says that there is no algorithm which works in general.

In one sense, this answer is very disappointing. An algorithm, if it existed, would enable us to settle all the outstanding questions of number theory in a routine way. On the other hand, it is comforting to know that mathematicians cannot be replaced by computers.

8.4 Gödel's First Incompleteness Theorem

Let us look more carefully at arithmetic, i.e. the set of all sentences of our formal language true in the standard interpretation. **BJ** Theorem 3 (page 176) tells us that this set is not decidable. The question arises as to whether there is some decidable set of sentences which, when taken as a set of axioms, yields arithmetic as its theory.

DEFINITION A set of sentences T is *axiomatizable* if there is a decidable set of sentences that can be taken as axioms for a system whose set of theorems is T.

Note the distinction between an axiomatizable theory and an axiom system. The former has a decidable set of axioms while the set of axioms of the latter may or may not be decidable.

Our question above, which this section is devoted to answering, can now be phrased as: is arithmetic axiomatizable? The fact that arithmetic is not decidable (**BJ** Theorem 3, page 176) does not rule out the possibility that this set is axiomatizable. After all, we know that the theory of the system Q is also not decidable (by Lemma 4 of **BJ** page 175) but it is axiomatizable, because Q has a set of axioms which is finite, and hence certainly decidable.

We also know that Q does not provide an axiomatization of arithmetic that is *complete* in the sense that all the sentences which are true in the standard interpretation are provable from its axioms. We discovered in *Unit* 7, Section 7.3, that there are such true sentences which are not theorems of Q. However, much stronger axiom systems for formal number theory are known. One such is discussed in the next section, and by 1930 considerable evidence had accumulated which pointed to its being a complete axiomatization. That is why Gödel's discovery that no such complete axiomatization exists came as a shock.

The system Q is an example of a theory which is not decidable: in other words, the gödel numbers of the theorems of Q do not form a recursive set. We begin by showing that the set of gödel numbers of the theorems of an axiom system does have a weaker but related property which we now explain.

DEFINITION A set θ of natural numbers is said to be *recursively enumerable* if it can be expressed in the form

$$\theta = \{f(0), f(1), f(2), \ldots\}$$

where f is a recursive function. In other words, a set is recursively enumerable if it is the range (set of values) of a recursive function. Note that our notation is not intended to imply that $f(0), f(1), f(2), \ldots$ are all distinct, so, as shown by the examples below, a recursively enumerable set can be finite. (It is also convenient to adopt the convention that the empty set counts as being recursively enumerable, but this will not concern us in this course.)

It follows that a set is recursively enumerable if and only if there is an algorithm for enumerating (or listing) its members. (There is a difference between a recursive set and a recursively enumerable set, as you will see later.)

EXAMPLE 1 The set $\{0\}$ is recursively enumerable. The zero function, z, is recursive and since for all n, $z(n) = 0$, the range of values taken by z consists of just the set $\{0\}$.

EXAMPLE 2 The set $\{0, 1, 4, 9, \ldots\}$ of perfect squares is recursively enumerable. The function f defined by $f(n) = n^2$ is recursive and its range of values is the set of perfect squares.

EXAMPLE 3 We give an informal argument to show that the set of numbers whose decimal representations occur as consecutive digits somewhere to the right of the decimal point in the decimal expansion of π is a recursively enumerable set. As $\pi = 3.14159\ldots$ this set includes, e.g. 5 (coming from the underlined part of $3.141\underline{5}9$) and 415 (coming from the underlined part of $3.14159\ldots$). There are well known series expansions which enable us to calculate the decimal expansion of π as far as we like. We can use this expansion to enumerate all finite sequences of digits which occur in this expansion after the decimal point, as follows.

Step 1 $\pi = 3.1\ldots$, thus the list begins $\{1, \ldots\}$.
Step 2 $\pi = 3.14\ldots$, thus we add to our list the sequences 4 and 14 so that it now is $\{1, 4, 14, \ldots\}$.
Step 3 $\pi = 3.141\ldots$, thus we add to our list the sequences 1, 41 and 141 so that it now is $\{1, 4, 14, 1, 41, 141, \ldots\}$ (illustrating again the point that the numbers in the list need not be distinct—here the sequence 1 appears more than once in the list).
and so on.

Since we have an algorithm for calculating the decimal expansion of π as far as we like, we have an algorithm which enables us to continue this list as far as we like. The set, θ, of all the numbers which occur in this list, is thus a recursively enumerable set. ☐

Notice that in Example 3 we gave only an informal argument to show that θ is a recursively enumerable set. We did not actually construct a recursive function whose range is θ, as would be required to give a completely rigorous proof. To avoid technical complications we shall stick to informal arguments in what follows.

From our point of view, the most important examples of recursively enumerable sets are given by the following theorem.

THEOREM 1 Let T be an axiom system for formal arithmetic with a decidable set of axioms. Then the gödel numbers of the theorems of T form a recursively enumerable set.

PROOF In *Unit 1*, Section 1.1, we showed that if we have an 'alphabet' consisting of a finite set of symbols, then the set of all finite strings of these symbols forms an enumerable set. Furthermore, there is an algorithm for enumerating it, namely, first list all the one-letter strings in alphabetical order, then list all the two-letter strings in alphabetical order, and so on.

In writing down formal proofs of our system we use only a finite set of symbols. These are the symbols of our formal language together with the symbols used to write the annotations (these are just letters of the English alphabet, the numerals $0, 1, 2, \ldots, 9$, and a few punctuation marks). So, as explained in the previous paragraph, there is an algorithm for listing all the finite strings of these symbols. Let the list generated by this algorithm be

$$S_0, S_1, S_2, \ldots. \qquad (*)$$

We stressed throughout *Units 5* and *6* that there is an algorithm for determining whether or not a purported proof really is a legitimate proof of our system, and, if so, on which assumption formulas the last line of the proof depends. Since T has a decidable set of axioms, there is an algorithm to determine whether or not a formula is an axiom of T. We can then use this algorithm to decide whether or not the assumptions on which the last line of a legitimate proof depends are all axioms of T.

It follows that we have an algorithm which enables us, as we generate the list $(*)$, to discard those strings which are not legitimate proofs whose last line depends on assumptions which are all axioms of T. Using this gives us an algorithm which generates a list

$$P_0, P_1, P_2, \dots \qquad (\dagger)$$

of all legitimate proofs of which the last line depends on assumptions which are all axioms of T. The formulas occurring on these last lines will therefore be all the theorems of T. Hence, if, as we generate the list (†), we calculate and list the gödel numbers of these formulas, we shall enumerate all the gödel numbers of the theorems of T. Since this enumeration is done by an algorithm, the set of all these gödel numbers is recursively enumerable. ■

As a consequence of Theorem 1, the set of gödel numbers of theorems of the system Q can be recursively enumerated, in contrast with the fact that this set is not recursive (the content of **BJ** Lemma 4, page 175). (Recursively enumerable sets are of intrinsic interest to logicians, and the above discussion gives a proof of the result that there are recursively enumerable sets which are not recursive.)

The moral to be stressed here is that the existence of an algorithm for enumerating the theorems of an axiom system does not imply the existence of an algorithm for deciding whether or not a formula is a theorem. 'Wait and see if it appears in the list' is not an algorithm for giving an answer to this question. If a formula is a theorem then it will appear at some stage of the enumeration and when it appears we will know that it is a theorem. But if it is not a theorem there will never be a stage when we can be sure of this. The fact that a formula has not so far appeared in the enumeration gives us no information about whether it will appear later.

There is, however, one special case where we can deduce that the theorems of an axiom system form a decidable set.

DEFINITION The theory T is said to be *complete* if for each sentence A, either $\vdash_T A$ or $\vdash_T -A$.

EXAMPLE 4 Arithmetic is a complete theory. By definition, arithmetic consists of all the sentences true under the standard interpretation. Given a sentence A, either A is true under this interpretation and hence is an axiom of, and thus a theorem of, arithmetic, or A is false under this interpretation, in which case $-A$ is true and hence likewise a theorem of arithmetic.

THEOREM 2 The theorems of a complete and consistent axiomatizable theory form a decidable set.

PROOF Let T be a complete and consistent axiomatizable theory. By Theorem 1, there is an algorithm for enumerating the theorems of T. This algorithm produces a never-ending list $A_0, A_1, A_2, \dots$ of all the theorems of T. We describe an algorithm for deciding whether an arbitrary sentence is a theorem of T.

Given a sentence A, we know that as T is complete, either A or $-A$ is a theorem of T. Thus after a finite number of steps of the enumeration of all the theorems of T, one of the sentences A, $-A$ must appear in the list as a theorem. There are two possibilities, depending on which of these sentences appears.

(i) The sentence that appears in the list is A, so that A is a theorem of T.

(ii) The sentence that appears in the list is $-A$. As T is consistent we know that we cannot prove both $-A$ and A as theorems of T, so that we know that A cannot subsequently appear in the list as a theorem of T. So in this case A is not a theorem of T.

In either case the algorithm terminates with the correct answer. ■

We now come to the main theorem of the course.

THEOREM **Gödel's First Incompleteness Theorem** (**BJ** Theorem 6, page 179)
There is no complete consistent axiomatizable extension of Q.

PROOF By Theorem 2 above, a complete and consistent axiomatizable extension of Q would have a decidable set of theorems, and this would contradict **BJ** Theorem 1, page 175. So no such complete, consistent axiomatizable extension of Q exists. ■

COROLLARY Arithmetic is not axiomatizable.

PROOF Arithmetic is a complete and consistent extension of Q. ■

READ BJ page 179, line 15 to page 180, line 6.

8.5 Peano Arithmetic

Gödel's First Incompleteness Theorem tells us that we cannot find a consistent decidable set of axioms from which all the true sentences of arithmetic can be derived as theorems. In Section 7.3 of *Unit 7* we gave examples of such sentences (i.e. ones true under the standard interpretation) that were not theorems of the system Q: these included such basic things as $\forall x\ x' \neq x$ and $\forall x\ \forall y\ (x + y) = (y + x)$. Q is thus a very weak system. In this section we shall look at a decidable set of axioms from which we can prove somewhat more of the sentences true in the standard interpretation.

If you have done any number theory, you will probably find it easy to see what is missing from the system Q which makes it such a weak system. One of the most common methods of proof in number theory is the method of *mathematical induction*, but there is nothing in the system Q which corresponds to this.

The principle of mathematical induction says that for each property of numbers, if 0 has this property, and if whenever n has the property, so also does $n + 1$, then every natural number has the property. (Since most number theorists regard the natural numbers as starting with 1, they would word this slightly differently, but in essentially the same way.)

The axiom system we are about to study incorporates this principle in the way explained in the next reading passage.

READ BJ page 182, line 1 to page 183, line 5.

NOTES (i) *Page 182, lines 5, 6*
We explain the significance of '*Elementary*' in the next note. The system is named after the Italian mathematician Giuseppe Peano (1858–1932) who first devised a system of axioms for number theory in 1889. These axioms were somewhat different from those which make up the system described by **BJ**. 'Hilbert–Bernays' is a reference to a very influential book *Grundlagen der Mathematik* (Foundations of Mathematics) published by David Hilbert (1862–1943) and Paul Bernays (1888–1977) in two volumes in 1934 and 1939.

(ii) *Page 182, lines 8–10*
The principle of mathematical induction, as we mentioned above, states: 'for each property of numbers ...'. In our language, properties of numbers are expressed by formulas, and as no symbol in our language expresses 'for each property', the only way we can represent this principle is by taking for each formula $A(x)$ the following new axiom

$$([\underset{\substack{\uparrow\\ \text{0 has the}\\ \text{property}}}{A(\mathbf{0})} \quad \& \quad \underset{\substack{\uparrow\\ \text{if } n \text{ has the property}\\ \text{so also has } n+1}}{\forall x\,(A(x) \rightarrow A(x'))}] \quad \rightarrow \quad \underset{\substack{\uparrow\\ \text{every natural}\\ \text{number has the}\\ \text{property}}}{\forall x\, A(x)}).$$

If the formula $A(x)$ has free occurrences of variables in it other than x, we turn the above formula into a sentence (i.e. a formula with no free variables) by prefixing universal quantifiers for all the free variables other than x which occur in $A(x)$.

The term 'elementary' used to describe the theory Z is a technical one which refers to the fact that in the language of Z the only quantifiers we have are those that refer to numbers. So in the language of Z we can express 'for every number' but not 'for every property of numbers'.

(iii) *Page 182, line 14*
By 'logical truths', **BJ** means 'provable using no assumptions'.

(iv) *Page 182, lines −14, −13*
Although the induction axioms form an infinite set (because we get one induction axiom for each formula $A(x)$), Z is an axiomatizable theory. It should not be difficult to see that there is an algorithm for deciding whether a formula of our language has the form of an induction axiom. Hence the axioms of Z form a decidable set.

(v) *Page 182, lines −12 to −4*
We do not need to refer to the 'notion of *proof*' given by **BJ** since we have already introduced our own system of formal proofs in *Units 5* and *6*. We do not have the time to explore the strength of the system Z, but the evidence that had accumulated by 1930 certainly suggested that it might very well be possible to derive all true sentences within the system Z. That is why it was such a shock when Gödel proved his First Incompleteness Theorem in 1931. ▲

Since any consistent axiomatizable theory T which extends Q is not complete, there will be at least one true sentence (i.e. a sentence true under the standard interpretation) which is not derivable in T. Gödel's original proof of his First Incompleteness Theorem gave a method for explicitly constructing such a sentence given the axioms of T. Our proof, although it used essentially the same ideas as Gödel's, missed out some of the technical complications and as a result we did not construct an explicit example of a true but underivable sentence. We shall give one such example in the next section.

8.6 An Answer to Hilbert's Question

You will remember that Hilbert's question is: *can the consistency of number theory be proved using only non-dubious principles of finitary reasoning?*

To be able to answer this question we need to make it precise in various ways. Firstly, by 'number theory' we shall mean, at least for the time being, the system Z of Peano arithmetic introduced in the previous section.

Up to now we have taken it for granted that Z is consistent because all the axioms of Z are true in the standard interpretation. We argued thus:

> Since all our logical rules are valid, all the theorems of Z will also be true under the standard interpretation. Given a sentence A, it cannot be that both A and $-A$ are true in the same interpretation. So A and $-A$ cannot both be theorems of Z. Hence Z is consistent.

This argument does not meet Hilbert's requirements since it takes for granted the standard interpretation of number theory, and the notion of 'truth' in relation to this interpretation. But the standard interpretation of number theory is an infinite mathematical object and arguments which assume that we have a clear notion of truth for such infinite structures go beyond 'non-dubious principles of finitary reasoning'. Indeed, Gödel's First Incompleteness Theorem shows that the properties of this standard interpretation cannot be completely captured by any explicitly given set of axioms of our language. Thus our grasp of the structure of the standard interpretation is not as straightforward a matter as might be suggested by the consistency proof for Z given above.

We can interpret Hilbert's requirement as seeking an axiom system with far fewer axioms than Z in which the consistency of Z can be derived. In other words, can we find an axiom system T such that Z is an extension of T and such that the sentence which expresses the consistency of Z is a theorem of T?

This immediately raises the question of how we can find a sentence of our formal language which expresses the consistency of Z. The following observation is helpful.

THEOREM 1 Z is consistent if and only if the sentence $\mathbf{0} = \mathbf{1}$ is not a theorem of Z.

PROOF Since $\mathbf{0} \neq \mathbf{1}$ is a theorem of Z (by Theorem 1 of *Unit* 7, Section 7.4), if Z is consistent then $\mathbf{0} = \mathbf{1}$ cannot be a theorem of Z. Conversely, if Z is not consistent then every formula is a theorem of Z (see the result on page 92 of Section 8.3). So if $\mathbf{0} = \mathbf{1}$ is not a theorem of Z, then Z must be consistent. ■

It follows that to express the consistency of Z all we need to express is 'the sentence $\mathbf{0} = \mathbf{1}$ is not a theorem of Z'. We could do this with the sentence $-B(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$, provided that $B(y)$ is a formula which expresses 'y is the gödel number of a theorem of Z'. The difficulty with this is that by **BJ** Lemma 3 (page 174) the set of gödel numbers of the theorems of Z is not definable in Z. However, it is possible to construct a formula Prov(y) which has the property

> $\mathrm{Prov}(\ulcorner A \urcorner)$ is true in the standard interpretation if and only if $\vdash_Z A$.

The construction of this formula Prov(y) involves a good deal of technical complication and so we shall not describe it. We can use this formula to restate Hilbert's question as follows:

> Is there an axiom system T such that Z is an extension of T and $\vdash_T -\mathrm{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$?

We show that the answer to this question is 'no' by proving that if Z is consistent then it is not even the case that $-\mathrm{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$ is a theorem of Z, and so it cannot be a theorem of a system which is weaker than Z.

Rather than prove this for the particular formula Prov(y), it is convenient to prove it in a more general setting which deals with any formula which might be considered to express the fact that formulas are derivable. **BJ** calls such formulas *provability predicates.*

READ BJ page 185, line 7 to page 186, line 5, omitting page 185, lines 17–21.

NOTE *Page 185, lines 14–16*

If we think of $B(\ulcorner A \urcorner)$ as expressing 'A is a theorem of T', these conditions say:

(i) if A is a theorem of T, then we can prove in T that A is a theorem of T;

(ii) we can prove in T that if $(A \rightarrow C)$ and A are theorems of T then C is also a theorem of T;

(iii) we can prove in T that if A is a theorem of T, then we can prove in T that A is a theorem of T. ▲

The theorem which finally enables us to answer Hilbert's question is a very ingenious application of Gödel's Diagonal Lemma. The individual steps in the proof are not difficult to follow; they use the Tautology Rule, which **BJ** does not mention explicitly. However, once again it is difficult to get a grasp of the proof as a whole. So you should not worry if you find it hard to see what is really going on. We shall not expect you to remember either the theorem which follows in the next reading passage or its proof. We have included it in your reading to indicate what is involved in answering Hilbert's question.

READ BJ page 186, line 11 to page 188, line 16.

It is a bit misleading to call **BJ** Corollary 2 Gödel's Second Incompleteness Theorem, which really asserts that $-\text{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$ is not a theorem of Z. To draw this conclusion from the Corollary we would need to prove that $\text{Prov}(y)$ is indeed a provability predicate for Z. The detailed verification of this is complicated and so we omit it. (In missing out this step of the argument we are in distinguished company: Gödel's original proof of the Second Incompleteness Theorem was different from the one given here, but Gödel also never published the proof in detail. He gave an outline of the proof in his 1931 paper and said that the details would be given in a subsequent paper; but this was never published. Following in this tradition, almost all the textbooks which give a proof of the Second Incompleteness Theorem just wave their hands at this stage of the argument.) So we just state without proof:

LEMMA $\text{Prov}(y)$ is a provability predicate for Z.

It is worth remarking that here we do really need a powerful system with lots of axioms like Z. Being a provability predicate for Z means that certain formulas (given in conditions (i), (ii), (iii) on **BJ** page 185) are provable, and for this to be the case, a system like Z is needed. The analogous lemma for a system with very few axioms like Q would not be true.

From this Lemma we can immediately deduce:

HEOREM 2 **Gödel's Second Incompleteness Theorem**
If Z is consistent then $-\text{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$ is not a theorem of Z.

We have included the hypothesis that Z is consistent in our statement of this theorem since its main interest stems from Hilbert's question, and this question has little significance if the consistency of Z is taken for granted. However, even in this case the theorem is interesting. If Z is consistent, $-\text{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$ is true under the standard interpretation, and so is an example of a true sentence which is not derivable in **Z**.

Now we are in a position to answer Hilbert's question. Taking together the results of Theorems 1 and 2 we can conclude that

if Z is consistent then we cannot prove in Z that it is consistent.

Thus the answer to Hilbert's question is 'no'.

Theorem 2 applies more generally to any system with a decidable set of axioms which is an extension of Z, but it is important to note that the formula $\text{Prov}(y)$ varies with the system being considered. If we want to prove Theorem 2 for the axiom system T we must take $\text{Prov}(y)$ to be the formula which expresses 'y is the gödel number of a theorem of T' and the exact form of this formula will depend on the system T.

There is a cassette-tape commentary on the results of this and the previous sections. You will find it convenient to have to hand the list of results in objective (iv) of Section 8.8 when you listen to the tape.

START THE TAPE WHEN YOU ARE READY

8.7 In Conclusion

We have ended this course with several negative results. The answer to both Leibniz's and Hilbert's questions is 'no'. However, our work in this course has positive aspects. The concepts we have had to study, especially these of recursive functions and axiom systems, have been very fruitful in many areas we have not had the space to discuss. For this reason we should like to make some suggestions for further reading.

The literature of mathematical logic is very large and is still growing. Unfortunately, no two books use exactly the same notation and nomenclature, so you have to be ready to get used to different ways of presenting the same material. We have selected a short list of books which most appeal to us. Most of them will be available only in specialist libraries, but should be available from any library via the Inter-library Loan scheme. Also, many of these books are expensive, so we recommend that you do not buy any of them until you have been able to borrow a copy and satisfy yourself that it suits your needs.

(a) Historical

William and Martha Kneale's *The Development of Logic* (Clarendon Press, Oxford, 1962) is a scholarly work with a great deal of material on Greek and medieval logic, G. T. Kneebone's *Mathematical Logic and the Foundations of Mathematics* (Van Nostrand, London, 1963) is more technical and has a greater emphasis on work done this century. Jean van Heijenoort's *From Frege to Gödel* (Harvard UP, 1967) is an edited collection of classical papers written on mathematical logic in the period from 1879 to 1931. A translation of the paper in which Gödel originally proved his incompleteness theorems is included. The papers in this volume make difficult reading, but there are many helpful editorial comments.

(b) Philosophical

We begin with two classical books, Gottlob Frege's *The Foundations of Arithmetic* (first published in 1884 and republished with an English translation by J. L. Austin by Blackwell, Oxford, revised edition 1953) and Bertrand Russell's *Introduction to Mathematical Philosophy* (Allen and Unwin, London, 1919). A valuable anthology of various writings on the philosophy of mathematics is *Philosophy of Mathematics: Selected Readings*, edited by Paul Benacerraf and Hilary Putnam (Blackwell, Oxford, 1964).

(c) Computability

There are now a large number of textbooks dealing with recursive functions but none uses exactly the same notation as **BJ**. One of the earliest of these textbooks is Martin Davis's *Computability and Unsolvability* (McGraw-Hill, New York, 1958) and one of the most recent is Nigel Cutland's *Computability: An Introduction to Recursive Function Theory* (Cambridge UP, 1980). J. Hartley Rogers' *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, New York, 1967) deals with the more advanced aspects of the theory.

(d) Basic Logic

E. J. Lemmon's *Beginning Logic* (Nelson, London, 1965) describes a formal system which is quite close to the one described in this course. The main difference is that our Tautology Rule is replaced by a number of different rules for the symbols &, $\vee$, $\rightarrow$, $-$. Among the many textbooks which develop mathematical logic as far as this course, we mention only A. G. Hamilton's *Logic for Mathematicians* (Cambridge UP, 1978).

(e) Advanced Logic

Two recently published books, which coincidently have the same title, provide general accounts of the directions in which mathematical logic has developed. These are John Bell and Moshe Machover's *A Course in Mathematical Logic* (North-Holland, Amsterdam, 1977) which is based on their London University MSc course, and Yu. I. Manin's *A Course in Mathematical Logic* (Springer-Verlag, New York, 1978).

(f) Fun

An unusual but entertaining book which looks at formal systems, self-reference, and Gödel's Incompleteness Theorems is Douglas R. Hofstadter's *Gödel, Escher, Bach: An Eternal Golden Braid* (Harvester Press, 1979; Penguin, 1980).

8.8 Objectives of Unit 8

We list those things on which we may set assessment questions to test your understanding of the unit. After working through the unit we feel you should be able to:

(i) calculate the gödel number of a formula and decode a number to get the formula it encodes—in an examination question you would be given the particular numbers assigned to the basic symbols, so there is no need to remember these;

(ii) write down and interpret the diagonalization of a given formula;

(iii) use the technique of Problems 1 and 2 of Section 8.2 to show that certain types of function are recursive;

(iv) explain the meaning of each of the following results (and thus of the words used to state them), and how they relate to Leibniz's and Hilbert's questions:

Results

(1) *Gödel's Diagonal Lemma* (**BJ** *Lemma 2, page 173*)
Let T be an axiom system in which the function diag is representable. Then given any formula $B(y)$ with one free variable y, we can find a corresponding sentence G such that

$$\vdash_T (G \leftrightarrow B(\ulcorner G \urcorner)).$$

(2) **BJ** *Lemma 3, page 174*
If T is a consistent extension of Q, then the set of gödel numbers of theorems of T is not definable in T.

(3) **BJ** *Theorem 1, page 175*
No consistent extension of Q is decidable.

(4) **BJ** *Lemma 4, page 175*
Q is not decidable.

(5) **BJ** *Theorem 3, page 176*
Arithmetic is not decidable.

(6) *Theorem 1 of Section 8.4*
Let T be an axiom system with a decidable set of axioms. Then the set of gödel numbers of the theorems of T is a recursively enumerable set.

(7) *Theorem 2 of Section 8.4*
The theorems of a complete and consistent axiomatizable theory form a decidable set.

(8) *Gödel's First Incompleteness Theorem* (**BJ** *Theorem 6, page 179*)
There is no complete consistent axiomatizable extension of Q.

(9) *Gödel's Second Incompleteness Theorem* (*Section 8.6*)
If Z is consistent then $-\mathrm{Prov}(\ulcorner \mathbf{0} = \mathbf{1} \urcorner)$ is not a theorem of Z.

(We shall not test your understanding of the proofs of these theorems.)

Index (Units 5 to 8)

Words

Adequacy Theorem, 62
antecedent, 24
arithmetic, 94
arithmetic symbols, 8
assumptions in force, 19
assumptions, rule of, 20
atomic formula, 13
axiom, 67
axiomatizable, 95
axiom system, 67

bi-implication, truth table for, 25
bound variable, 41

characteristic function of a set, 93
complete, 97
conclusion, 23
conditional proof, rule of, 32
conjunction, truth table for, 23
connective, 8
consequence
  logical, 22
  tautological, 30
consequent, 24
consistent, 92
contingent, 67
contradiction, 57
Correctness Theorem, 62

decidable set of sentences, **BJ** 174
definable set of numbers, **BJ** 173
diagonalization of a formula, **BJ** 172
Diagonal Lemma, 91, **BJ** 173
disjunction, truth table for, 24
domain, 40, 66
dummy variable, 41

Elementary Peano Arithmetic, **BJ** 182
existential hypothesis rule, 55
existential quantifier, 8
existential quantifier introduction rule, 53
extension of a theory, **BJ** 174

formal proof, 20
formula, 14
  atomic, 13
freely substituted, 46
free variable, 41
function
  characteristic, 93
  representable in Q, 80
  representable in T, 82

gödel number, **BJ** 170, 171
Gödel's First Incompleteness Theorem, 98, **BJ**
Gödel's Second Incompleteness Theorem, 101

identity introduction rule, 59
identity symbol, 8
implication, truth table for, 25
independent, 77
induction
  axioms for, **BJ** 182
  mathematical, 98
interpretation of $\forall$, $\exists$, 66
interpretation
  non-standard, 66, 75
  standard, 66

language, basic symbols of, 8
leading connective occurrence, 16
leading symbol occurrence, 36
length function, 90, **BJ** 172
Löb's Theorem, **BJ** 187
logical consequence, 22
logically valid, 22
logical symbols, 8

mathematical induction, 98

negation, truth table for, 24
non-standard interpretation, 66, 75

occurrence, 42

Peano Arithmetic, Elementary, 100, **BJ** 182
premise, 23
proof
  conditional, 32
  formal, 20
provability predicate, **BJ** 185

Q, axioms of the system, 68
quantifier
  existential, 8
  universal, 8

recursively enumerable set, 95
recursive set, 93
Representability Theorem, 82
representable function
  in Q, 80
  in T, 82
represents, 80

rules
  assumptions (Ass), 20
  conditional proof (CP), 32
  existential hypothesis (EH), 55
  existential quantifier introduction (EI), 53
  identity introduction (II), 59
  substitution (SR), 60
  tautology (Taut), 31
  universal quantifier elimination (UE), 47
  universal quantifier introduction (UI), 48

satisfiable, 92
sentence, 43
standard interpretation, 66
subformula, 16
substituted, freely, 46
substitution rule, 60
symbol
  arithmetic, 8
  basic, 8
  logical, 8
system, axiom, 67

tautological consequence, 30
tautology, 29
tautology rule, 31
term, 10
theorem, 67
theory, 67, 91
truth table
  for &, 23
  for $-$, 24
  for $\vee$, 24
  for $\rightarrow$, 25
  for $\leftrightarrow$, 25
truth value, 23
turnstile, 50

universal quantifier, 8
universal quantifier elimination rule, 47
universal quantifier introduction rule, 48
universe of discourse, 40

valid, logically, 22
variable, 8

Z, axioms for, **BJ** 182

Function Names and Special Symbols

Ass 20
CP 32
conj 90
diag **BJ** 172
EH 55
EI 53
gn **BJ** 171
II 59
lh 90, **BJ** 172
n 69
$\mathcal{N}$ **BJ** 172
N^* 75
N^{**} 75
neg 90
num **BJ** 172
Prov (y) 100
Q, $Q1$–$Q7$ 68
SR 60
Taut 31
UE 47
UI 48
Z **BJ** 182
— 8, 24
& 8, 23
$\vee$ 8, 24
$\rightarrow$ 8, 24
$\leftrightarrow$ 8, 25
$=$ 8
$\neq$ 69
$\forall$ 8
$\exists$ 8
$\mathbf{0}$ 8
$+$ 8
$'$ 8
$\cdot$ 8
$x_{''''}$ 8
0 23
1 23
$\phi(\tau/x)$ 45
$\vdash$ 50
$\vdash_T$ 69
$\phi(\mathbf{p}_1, \ldots, \mathbf{p}_n)$ 79
$\ulcorner A \urcorner$ **BJ** 172
μm 90
$\notin$ 92
$\not R$ 92
$*$ **BJ** 172

0 335 14013 0